Malabika Roy · Saikat Sinha Roy *Editors*

International Trade and International Finance

Explorations of Contemporary Issues

Springer

International Trade and International Finance

Malabika Roy · Saikat Sinha Roy
Editors

International Trade and International Finance

Explorations of Contemporary Issues

Springer

Editors
Malabika Roy
Department of Economics
Jadavpur University
Jadavpur, Kolkata, West Bengal
India

Saikat Sinha Roy
Department of Economics
Jadavpur University
Jadavpur, Kolkata, West Bengal
India

ISBN 978-81-322-2795-3 ISBN 978-81-322-2797-7 (eBook)
DOI 10.1007/978-81-322-2797-7

Library of Congress Control Number: 2016937967

© Springer India 2016
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer (India) Pvt. Ltd.

Foreword

The present volume is a collection of articles on various aspects of international trade and international finance, written by a selected group of researchers who have specialized in the respective fields. The planning of this volume was mooted at the meeting of advisory committee of the Centre for Advanced Studies (CAS) of the Department of Economics, Jadavpur University. Since CAS has sponsored a variety of workshops and seminars and hosted visiting fellows, the external experts of the CAS, Profs. Dilip Nachane (former Director, Indira Gandhi Institute of Development Research, Mumbai) and Ramprasad Sengupta (formerly of Jawaharlal Nehru University, New Delhi) suggested that we collate important papers in two volumes to be contributed by the visitors as well as Jadavpur's own faculty. The reason for planning two volumes is to accommodate the five thrust areas we have in the CAS, namely International Trade, Finance, Resource and Environment, West Bengal Economy and Public Policy related to social sector. The idea is to produce a monograph which will help the readers with some survey papers along with some state-of-the-art research in relevant fields. The present volume *International Trade and International Finance: Explorations of Contemporary Issues* edited by my two junior colleagues, Malabika Roy and Saikat Sinha Roy, who are themselves specialists in the subject matters included in the book, has a very well-thought collection of topics covered under the purview of the book. The papers selected will be of great help for the teachers and researchers in the subject on the one hand and on the other, it would certainly be a handy reference for policy planners and practitioners. I must thank the efforts of the editors who spared their valuable time to make the volume useful and interesting. I have faith that the purpose of publishing this volume will be totally fulfilled and we will be having increasing returns to knowledge after reading the volume.

Ajitava Raychaudhuri
Professor and Coordinator, Centre for Advanced Studies
Department of Economics, Jadavpur University, Kolkata, India

Contents

Introduction

International trade and international finance have always been important areas of research in economics. With globalization, the nature, importance, and scope of the subjects have changed and the horizon has expanded manifold. Instead of treating international trade and international finance as two disjoint areas of study, the present volume brings together a collection of essays from both the fields, sometimes overlapping across the two areas. The volume, while focusing on the recent developments and frontiers of research in international trade and international finance, also emphasizes the inherent integrated nature of the two subjects.

Theory and empirics in international trade is founded on microeconomic principles highlighting the gains from trade through specialization and exchange as a country moves from autarky to free trade. The traditional theoretical notion of comparative advantage evolved to include nuances of imperfect competition and product differentiation. This has again recently given way to "New-New Trade Theory", which centres on the concept of heterogeneity of firms. This latest development in trade theory is an important direction of trade research especially during globalization. With globalization, new sectors have emerged important and the modes of trade considerably vary from that in the past. Further, new institutions, both global and domestic, have emerged and have significant implications for trade. In this volume, the theoretical and empirical papers included deal with recent advances in the subject, emergence of new sectors and institutions in shaping up trade during globalization.

International finance primarily developed from macroeconomics to address the same issues about equilibrium level of income, growth of output and employment and effects of monetary and fiscal policies—but in an international context, when more than one economy are interacting with each other in the sphere of trade and finance involving more than one currencies. It is not surprising that traditionally, the issues addressed in international finance were balance of payments and related policies, foreign exchange market and foreign exchange management, global capital markets and cross-border flow of funds, international financial systems and their management. However, globalization brought about not only an integration of financial markets but also an integration of financial institutions. So in the present volume we have included essays that address issues related to international financial institutions along with essays dealing with traditional issues on foreign exchange markets and international financial markets.

The present volume contains 28 essays divided into seven thematic parts. Rather than dividing the parts in line of international trade and international finance, we have divided the parts according to thematic uniformity thus establishing a close link between international trade and international finance. However, the chapters included in Parts I–V focus more on issues related to international trade, whereas chapters included in Parts V–VII lean more towards issues related to international finance.

Part I covers works on recent developments in international trade theory and empirics. Chapter "The "New-New" Trade Theory: A Review of the Literature" by Priya Ranjan and Jibonayan Raychaudhuri is an account of the developments and progress in "New-new" trade theory models and empirical research centred on the seminal work of Melitz (2003). This chapter also discusses the policy implications and welfare implications of trade liberalization in this context. In Chapter "Time Zones and FDI with Heterogeneous Firms", Toru Kikuchi, Sugata Marjit and Biswajit Mandal develop a model on the role of FDI in the context of heterogeneous firms situated in different time zones. It is shown that firms undertaking FDI have higher productivity than non-FDI firms, and the foreign subsidiaries of high-productivity firms serve the home market. In Chapter "MNEs and Export Spillovers: A Firm-Level Analysis of Indian Manufacturing Industries", in an empirical analysis, Maitri Ghosh looks into the role of firm heterogeneity in export spillovers in the presence of FDI. It is found that firm heterogeneity, measured in terms of productivity and sunk cost, is critical to explaining export spillovers in the presence of multinationals. In contrast to the other chapters in this part, Chapter "IPR Regulatory Policy, Commercial Piracy and Entry Modes of MNC: A Theoretical Analysis" by Nilanjana Biswas (Mitra) and Tanmoyee Banerjee (Chatterjee), relates IPR regulation regime chosen by LDC government to the mode of entry of the MNC: fragmentation or full technology transfer, in the presence of commercial piracy. It is found that entry in this case depends on the transport cost and monitoring.

The second part on international trade and institutions covers issues like trade openness and the size of government, issues related to bilateral, regional and multilateral trade. In Chapter "International Trade and the Size of the Government", using a theoretical framework Rajat Acharyya shows that under certain conditions pertaining to the non-traded public good, trade liberalization in terms of tariff reduction necessarily increases absolute size of the government. The relative size of the government expands when the value of the price elasticity of the public good is small though not less than unity. Corruption in trading countries at the importer level as well as exporter level plays an important role in bilateral trade. In Chapter "The Effects of Corruption on Trade Flows: A Disaggregated Analysis", Subhayu Bandopadhyay and Suryadipta Roy investigate the impact of importer level and exporter level corruption on bilateral exports for 27 sectors of 100 countries during 1984–2004. The other chapters in this part deal with regional trading blocs or issues related to WTO. Using an oligopolistic framework in Chapter "Enlargement Decisions of Regional Trading Blocs", Sunandan Ghosh shows the possibilities of and nature of equilibrium in the expansion or consolidation of regional trading blocs in the presence of technology and market asymmetries between countries. The conclusions are of importance for emerging market economies, which are a part of an existing regional trading arrangement seeking expansion or consolidation. In Chapter "Deal Breaker or the Protector of Interests of Developing Countries? India's Negotiating Stance in WTO", Parthapratim Pal, delineates India's engagements in Doha Round of trade talks in the light of the experience with the WTO regime and changes in global trade including proliferation of RTAs along with rising global commodity prices. The last chapter in this part, an empirical paper by Saikat Sinha Roy and Pradyut Kumar Pyne on "Is WTO Governed Trade Regime Sufficient for Export Growth?", highlights that a WTO-governed trade regime is not sufficient for export growth across countries.

In the third part, which focuses on trade and development, there are theoretical as well as empirical papers. Chapters "Export Performance in Textile and Garments with China as a Competitor: An Analysis of India's Situation from the Perspective of Structure-Conduct-Performance Paradigm" and "Impact of Trade Liberalization on Indian Textile Firms: A Panel Analysis" both deal with different aspects of trade in India's textiles sector, a very important sector in terms of output, wide range of technology used, foreign exchange earnings and employment. In Chapter "Export Performance in Textile and Garments with China as a Competitor: An Analysis of India's Situation from the Perspective of Structure-Conduct-Performance Paradigm", Sarmila Banerjee, Sudeshna Chattopadhyay and Kausik Lahiri analyse textiles exports from India as compared to that from China, which followed a more aggressive approach in terms product and market diversification. The chapter relates such international price and non-price competitiveness to policies adopted in these countries. On the other hand, in Chapter "Impact of Trade Liberalization on Indian Textile Firms: A Panel Analysis" Subhadip Mukherjee and Rupa Chanda show the impact of trade liberalization on improvements in profitability, sales and import of raw materials. In this chapter, the effect of trade liberalization is found to be stronger through the import sourcing channel. The impact of merchandise trade on income inequality has been significant. In Chapter "Trade, Infrastructure and Income Inequality in Selected Asian Countries: An Empirical Analysis", Ajitava Raychaudhuri and Prabir De show that trade openness and infrastructure have significant impact on income inequality across countries in the Asia-Pacific region, and country-specific factors turn out to be important determinants of trade openness and income inequality. The chapter also establishes persistence in trade openness and income inequality across these countries.

The other chapters in this part deal with issues related to services trade, labour movements and development. Chapter "A Theoretical Model of Trade, Quality of Health Services and Signalling" by Kausik Gupta and Tonmoy Chatterjee build a theoretical model of services trade, health services in particular, and the role of quality signalling therein with respect to southern countries. The chapter arrives at optimum values of health quality and prices of health quality through a two-stage dynamic game. In Chapter "Smuggling and Trafficking of Workers: A Brief Review and Analysis of the Economics of Illegal Migration" Saibal Kar surveys the issue of economics of illegal international migration in general and smuggling or trafficking of workers. The chapter discusses a policy to lower exploitation from illegal labour migration. The last two chapters in this part deal with the incidence of child labour in the context of an open developing economy. Chapter "Impact of Trade Restriction on Child Labour Supply and The Role of Parents' Utility Function: A Two Sector General Equilibrium Analysis" by Biswajit Chatterjee and Runa Ray shows that non-trade policies are effective to deal with the incidence of child labour and trade policies remain ineffective.

In Part IV the main focus is on flow of foreign investment, i.e. FDI and FII and the role of multinationals. Chapter "The Determinants of Foreign Direct Investment: An Analytical Survey" by Kunal Sen and Chaitali Sinha surveys the literature on the factors determining the location decision of FDI. It is found that economies of scale, management skill and innovative product technologies are the major determinants of location decision of MNCs along with regulatory policies of the government. The chapter analyses the locational decisions of the southern MNCs as against those in the advanced market economies in the context of changing FDI flows in recent years. Both Chapters "Foreign Direct Investment, Capital Formation and Growth" and "Foreign Direct Investment and Macroeconomic Indicators in India: A Causality Analysis" concentrate on foreign direct investment. In Chapter "Foreign Direct Investment, Capital Formation, and Growth" Prabirjit Sarkar examines the relationship between growth of fixed capital formation and direct foreign investment (as % of GDP) in a panel of 61 countries covering a time period of 1980–2006 using alternative methodologies. In Chapter "Foreign Direct Investment and Macroeconomic Indicators in India: A Causality Analysis", Jaydip Mukherjee and Basabi Bhattacharya analyse the pattern of movement of external capital flows to Indian economy in terms of foreign direct investment (FDI) and the probable impact of macroeconomic indicators, viz. real GDP growth, call money rate, US dollar exchange rate, inflation, T-bill rate, trade openness and Dow Jones Index value on the financial and overall performance of the economy from the period 1997–1998 to 2013–2014.

Part V contains four papers each focusing on different aspects of globalization and functioning of financial markets. In Chapter "Exploratory Study of Select Commodity and Equity Indices Around the Meltdown of 2008" Diganta Mukherjee and Arnab Mallik conduct an exploratory study of commodity market performance in and around 2007–2008, when the world was hit by the meltdown resulting from subprime crisis. Their primary focus is to look into the nature of price movement and volatility some of the key base metals have been showing, mainly on the London Metal Exchange and MCX. They also attempt to establish patterns of movement in some of the popular equity indices and establish which one between the equity commodities fared well during the period of negative market sentiment in the years of subprime crisis. In Chapter "An Empirical Investigation of Volatility Clustering, Volatility Spillover and Persistence From USA to Two Emerging Economies India and China", Ayanangshu Sarkar and Malabika Roy examine the pattern of volatility in the Indian and Chinese stock market during 2006–2011 in terms of its time-varying nature, presence of certain characteristics such as volatility clustering and whether there exists any 'spillover effect' between the domestic and the US stock markets. They also estimate the persistence of shock in terms of half-life in each subperiod of study. In Chapter "Imbalances, Local and Global, and Policy Challenges in the Post Crisis World", Soumyen Sikdar addresses some interesting questions on correct policy choices in the face of the economic slowdown that India and China have experienced as a result of subprime crisis. Some very pertinent questions are: What are the major policy failures that allowed the catastrophe to happen? What type of reforms will prevent a recurrence? Should China, India and other developing economies of Asia and Africa make systematic efforts to 'decouple' from the developed countries and work towards greater integration among them? If yes, then how? The chapter attempts to suggest answers to these questions, after examining the impact on the Indian economy and the role of the Indian policymakers during the time of trouble. Future prospects of India and China, the two Asian giants, receive particular attention. In Chapter "Testing Non-linearity in Emerging and Developed Markets" Basabi Bhattacharya and Koushik Guhathakurata explore the possibility of non-linearity in selected stock markets of the world.

In Part VI we have included three papers all dealing with foreign exchange market, which is a major area of study in international finance. Chapter "Foreign Exchange Markets, Intervention and Exchange Rate Regimes" is a survey of structure of foreign exchange market. In this chapter Ashima Goyal describes the institutional features of FX markets, with special emphasis on the process of liberalization and deepening in Indian FX markets, in the context of integration of currency markets with financial markets and of large international capital flows. Chapter "Global Foreign Exchange Market: A Crisis Analysis" is an exploration of the global foreign exchange market dynamics around the significant financial meltdowns in the past. In this chapter Gagari Chakraborty studies the factors influencing the forex market movements over the last 20 years. She further enquires whether and how the sensitivity of forex market changes, following the changes in the chosen real and financial variables in times of such crises. Finally, in Chapter "The Impossible Trinity: Where Does India stand?", Rajeswari Sengupta addresses the dilemma of "impossible trinity". This chapter again is basically a survey chapter. In this chapter, she presents a comprehensive overview of a few empirical studies that have explored the issue of trilemma in the Indian context. Based on these studies she analyses how Indian policymakers have dealt with the various trade-offs while managing the trilemma over the last two decades and also draw relevant policy conclusions.

Part VII brings together four studies on financial institutions. In Chapter "Guaranty Funds and Moral Hazard in The Insurance Industry: A Theoretical Perspective", J.M. Gandar, Sumon Mazumdar and C.W. Sealey develop a model of the guaranty fund insurance company relationship under moral hazard, and examine the nature of adverse incentives in this setting. They also devise workable mechanisms that alleviate the moral hazard problem. Rama Seth and Kamran

Quddus, in Chapter "Performance of Aggregate Portfolios of Equity Mutual Funds: Skill or Luck?" present some descriptive statistics on mutual funds contrasting the two countries: USA and India. Then they evaluate the performance of mutual funds in an emerging market such as India, borrowing a methodology extensively used in asset pricing literature. Next using bootstrap simulations, they analyse the persistence of fund returns, distinguishing skill from luck. The next two chapters both deal with different aspects of international banking. Chapter "Foreign Bank Presence and Financial Development in Emerging Market and Developing Economies: An Empirical Investigation" by Sasidaran G. sets out to explore the empirical determinants of foreign bank entry in emerging and developing economies (EMDEs). Using panel data for over 100 EMDEs, this chapter contributes to the literature by throwing light on understanding the motives of foreign bank entry to EMDEs which remains a relatively under-researched topic in the literature. In Chapter "Banks, Financial Derivatives, and Crises: A Fourth-Generation Model", taking off from the third generation open economy model of financial crisis, Romar Correa examines the investment plans of domestic entrepreneurs supported by banks. He models outcomes consequent on changes in capital movements and the possibility of multiple equilibria.

A unique feature of the proposed volume is that it unravels some new issues in addition to re-examining certain old issues in a new perspective and it covers wide ranging issues with an emphasis on policy. The book covers issues mostly related to emerging market economies, which has increasingly assumed importance in the context of globalization. The book contains some survey papers covering the frontiers of current knowledge on important themes like recent developments in trade theory and empirics, foreign exchange market, interrelation and interaction between international trade and international finance. The book, thus, will be of immense use for advanced undergraduate and graduate teaching as well as for research. We expect the book to substantially contribute to the growing literature on issues related to trade and international finance in emerging market economies and extend the frontiers of knowledge. The editors are grateful to the Centre for Advanced Studies, Department of Economics, Jadavpur University, Kolkata, and in particular to Ajitava Raychaudhuri and Basabi Bhattacharyy for entrusting them with the job. The editors are extremely grateful to the authors as well as the reviewers of the papers for their carrying out their respective duty with responsibility. The volume would not have seen the light of the day without the extreme cooperation of the editorial team of Sringer India.

Part I
Recent Developments in Trade Theory and Empirics

The "New-New" Trade Theory: A Review of the Literature

Priya Ranjan and Jibonayan Raychaudhuri

Abstract We review the literature on the so-called "new-new" trade theory models starting with the pioneering work by Melitz (Econometrica, 71(6):1695–1725, 2003). We review some of the empirical work that motivated the development of these "new-new" trade theory models. We provide a survey of the theoretical literature on the "new-new" trade theory models and give a short account of the recent empirical work in this area. We also discuss policy implications and welfare implications of trade liberalizations in the context of this framework.

1 Introduction and Motivation

The development of trade theory has historically been driven by the discovery of stylized facts in data which required an explanation. "Old" trade theory, confronted by the empirical results of Grubel and Lloyd (1975) was found wanting in an explanation of these facts.[1] Old trade theory predicted a lot of "dissimilar-dissimilar" trade–trade in very different types of goods between countries with

[1]By "Old" trade theory we mean the traditional comparative advantage-based models of trade. Comparative advantage between countries emerges either from differences in labour productivity due to technology (the Ricardian model) or from differences in natural resources (the Heckscher–Ohlin Model). Comparative advantage predicts that under perfect competition, if trade is free from restrictions, countries produce and export those goods that they can make at a relatively (compared with other countries) lower opportunity cost.

P. Ranjan
Department of Economics, University of California, Irvine, USA

J. Raychaudhuri (✉)
Department of Economics, University of East Anglia, Norwich, UK
e-mail: j.raychaudhuri@uea.ac.uk

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_1

different resource endowments or technology.[2] However, the empirical evidence found by Grubel and Lloyd showed the prevalence of "similar-similar" trade (or *intra-industry* trade)–trade in the same types of goods between countries that are very similar (see Krugman 2008, p. 336). Krugman (1979, 1980) pioneered the development of "New" trade theory to explain these intra-industry trade flows.[3] The scope of this new trade theory was expanded to include the earlier traditional framework, in the synthesis of new and old trade theory by Helpman and Krugman (1985). However, this new trade theory, in turn, was not found to be broad enough to accommodate a number of stylized facts, observed in firm micro-data, around the 1990s. To explain these facts the (so-called) "new-new" trade theory was developed, pioneered by Melitz (2003).

Broadly these facts, as summarized in Bernard et al. (2007), are the following[4]:

(1) *Exporting is a rare activity*: In any industry only a small fraction of firms export. Also, exporters sell most of their output domestically.
(2) *Exporters are "better" than non-exporters*: Exporters are bigger and more productive than non-exporters (measured using either labour or total factor productivity) and they pay higher wages than non-exporters.

---

[2]Old trade theory was partially successful in explaining inter-industry trade patterns among countries. In a notable early study Bowen et al. (1987) use cross-sectional data for 1967 to test the "factor content" version of the Heckscher–Ohlin model. They use 27 countries, 12 factors of production and several goods in their analysis. For each country in their dataset, they compute the country's share of the world endowment of each factor and the country's share of world income. The Heckscher–Ohlin model predicts that a country should "export" factors (embodied in goods produced) in which the factor share of the country is higher than the share of the country's income and vice versa. The authors find (using simple sign tests) that the "… proportion of sign matches exceeds 50 % for 18 countries, and exceeds 90 % for five countries (Greece, Hong Kong, Ireland, Mexico, and the UK). However, the proportion of sign matches is below 70 % for 19 of the 27 countries." (Bowen et al. 1987, p. 796). In other words, the trade flows have the predicted sign (direction) less than 70 % of the time for most countries in the sample. Later empirical work has had more success by relaxing some of the assumptions of the Heckscher–Ohlin model. For example, Trefler (1995) allows for differences in technology between nations and a home bias in consumption. However, old trade theory assumed away completely the presence of intra-industry trade.

[3]The common thread in Krugman's models is a simple general equilibrium model of imperfect competition in a differentiated product industry. The basic assumption is that on the demand side consumers derive utility by consuming different varieties of a good. This is known as consumers' "love of variety" for differentiated goods (consumers' preferences are given by a CES utility function). On the production side, there is increasing returns to scale which makes it unprofitable for a single firm to produce all varieties. Under trade, consumers' driven by love of variety, demand home and foreign varieties of a good. Domestic and foreign firms (identical firms located either at home or abroad) specialize in producing and exporting only one variety of the differentiated good to take advantage of specialization from increasing returns. So a combination of love of variety and economies of scale ensures that large volumes of different varieties of a good are produced by different firms and then traded across countries. Krugman (1980) is a generalization of Krugman (1979).

[4]Categorization of the empirical evidence into these three broad heads and their subsequent discussion follows the approach outlined in Bernard et al. (2007).

(3) *Trade liberalization increases the average productivity level in an industry*: Following a period of trade liberalization average (weighted) productivity in a industry increases.

Since this empirical evidence played a crucial role in motivating the work by Melitz (2003) and the subsequent theoretical literature, we now take a closer look at the empirical evidence for each of the aforementioned points.

2 Stylized Facts in Need of an Explanation

First, let us consider point (1). In their survey article, Bernard et al (2007) report how rare exporting is. According to these authors, only 4 % of the 5.5 million U.S. firms were exporters in 2000. Among these 4 % of firms, the top 10 % accounted for 96 % of total U.S. exports.[5] Since in new trade theory all firms are identical by assumption, it cannot explain why some firms export and others do not.

Point (2) has been documented in very early studies by Bernard and Jensen (1995, pp. 81–87, 1999) (and reported in the survey by Bernard et al. (2007)).[6] They find that exporters outperformed non-exporters along a number of firm-specific attributes. Specifically, exporters are larger, more productive, more capital-intensive, more skilled-labour intensive and they also pay higher wages. For example, exporters are 119 % larger (measured in terms of employment), 26 % more productive (measured in terms of value added per worker), 32 % more capital-intensive (measured in terms of capital per worker), 19 % more skill-intensive (measured in terms of skill per worker) and pay 17 % higher wages than non-exporters.[7] Evidence on the better performance of exporting firms has been documented for other countries as well.[8] This feature—of the superior performance of exporters relative to non-exporters—regularly observed in data, is known as the "export premia" in the literature. Empirical studies have focussed particular attention on why these "export premia" arise. The critical question (from an academic as well as from a policy perspective) here is whether these premia arise

[5]See Bernard et al. (2007, pp. 108–109).

[6]One of the most comprehensive early studies is the work by Bernard and Jensen (1999) who use firm- and plant-level longitudinal data (from 1984–1992) from the Longitudinal Research Database (LRD) of the Bureau of the Census for US manufacturing firms. This database has detailed information on a number of firm and/or plant characteristics over a range of industries spanning U.S. manufacturing.

[7]See Bernard et al. (2007, Table 3 on p. 110).

[8]For example, Aw et al. (2000) report exporter premia for firms in China and Korea and Mayer and Ottaviano (2008) report premia for both exporters as well as firms engaged in FDI for firms in Germany, Hungary, Italy, Norway and the UK. Other important studies include Aw and Hwang (1995), Clerides et al. (1998).

because exporting is a difficult economic activity and so only the most productive firms are up to the task (thus a selection effect is at work) or is it because exporters learn and grow more efficient by exporting over time (thus a learning effect is at work). Most empirical results showed evidence for a selection effect, so a firm has to be highly productive before it can break into the export market.[9] By assumption, all firms in the Helpman–Krugman model are identical and so this theory again cannot explain why this feature is observed in the data.

Point (3), the increase in aggregate productivity post-trade liberalization, observed in recent empirical studies, is perhaps the most puzzling. The most important of these empirical studies is Pavcnik (2002) who considers the effect of Chile's trade liberalization, between the years 1979 and 1986, on productivity. Chile eliminated its non-tariff barriers, reduced its tariff rates and maintained a strong commitment to free trade in this period.[10] Also, the trade liberalization episode was accompanied by the exit of a number of plants.[11] Pavcnik uses a two-step method in her estimation. In the first step, she uses a semi-parametric method to obtain consistent estimates of the production function and then uses this estimated production function to calculate productivity.[12] In the second step, she measures the impact of trade liberalization on the estimated productivity from the first step.[13] Pavcnik uses (continuous) longitudinal data on Chilean plants for 7 years from 1979 to 1986. These plants are from eight manufacturing industries (two or three-digit ISIC industry-level) which are aggregated for convenience into import-competing, export-oriented and non-traded groups. To measure aggregate (industry-level) productivity she constructs a weighted productivity measure with weights proportional to the plant's output share in the industry in a year. She finds that aggregated weighted productivity increased in six out of the eight industries considered, over her sample period. Moreover, most of this improvement comes from resource and market share reallocation from less productive to more

[9]Van Biesebroeck (2005) is a notable exception and finds some evidence of learning by South African exporting firms. Also, Aw et al. (2000) find some evidence of learning by Korean firms. Ranjan and Raychaudhuri (2011) also find some evidence of learning by Indian firms.

[10]Tariff rates were reduced by more than 100 % during 1974 to a uniform ad-valorem tariff of 10 % in 1979. Trade liberalization continued throughout the 1980s except for a brief period from 1983–1984 when tariffs were raised to a 35 % uniform rate in response to a recession but were then reduced to an uniform rate of 20 % by 1985 (Pavcnik 2002, pp. 246–247).

[11]This suggests that the effects of plant liquidation need to be accounted for in the estimation process.

[12]This measure is plant specific and time varying.

[13]Pavnick's methodology deserves mention because it corrects for the simultaneity bias and the selection bias in measuring productivity. Simultaneity bias arises because a firm's productivity (known privately only by the firm) influences its choice of inputs in production leading to biased measures of productivity using just ordinary least squares regression (to estimate the production function). Selection bias arises because only those firms that except positive profits continue to produce and are therefore included in the estimation sample. Pavnick's semi-parametric estimation technique based on the method by Olley and Pakes (1996) controls for both of these biases.

productive firms. Overall, productivity increase is about 19 % after liberalization (over the seven-year period). Most of this gain (about 12.7 %) is due to reallocation of resources from low-productivity firms to high-productivity firms. Also, most of the reallocation changes occur in import-competing sectors and least in the non-traded sectors (see p. 262 and Table 3 in Pavcnik (2002)). In Chile’s case, trade liberalization was accompanied by reforms in other sectors of the economy which might have biased Pavcnik’s results. The evidence is much more convincing in Trefler (2004) who studies the effect of CUSFTA (Canada-United States Free Trade Agreement, 1988) on productivity in Canada.[14] Using data on Canadian industries from 1989–1996, Trefler obtains results similar to Pavcnik (2002). The free trade agreement leads to an overall increase in labour productivity (not total factor productivity as in Pavcnik (2002)) in Canada’s manufacturing sector by about 6 %. Once again, the sectors most impacted are the export-oriented group of industries (14 % increase in labour productivity) and the import-competing group of industries (15 % increase in labour productivity). Since the Helpman–Krugman model ignores firm heterogeneity by assuming identical firms, it cannot provide a reasonable explanation for the increase in average industry productivity following episodes of trade liberalization.

3 The Melitz Model and How It Explains the Stylized Facts

A number of stylized facts, therefore, remain unexplained by new trade theory. Melitz (2003) uses these stylized facts to motivate a model of trade where two key elements—firm heterogeneity in productivity and a fixed cost of entering export markets—determines the number and the type of firms that become exporters and the gains from trade. Melitz combines elements from both trade theory as well as industrial organization theory in his framework. For the trade part, Melitz builds on Krugman (1980), while for the industrial organization part he incorporates the dynamic industry equilibrium, set out in Hopenhayn (1992a, b), in his model.[15] We discuss the Melitz model in detail below since it explains a number of the stylized facts noted earlier and also because this work pioneered the burgeoning literature on heterogeneous firms. We first discuss the model under autarky and then look at the effects of trade in this model.

[14] Unlike Pavcnik (2002), Trefler (2004) is fortunate in having to consider only the (uncontaminated) effect of a preferential trade agreement between U.S. and Canada.

[15] Melitz mentions that his work “…draws heavily from Hopenhayn’s (1992a, b) work to explain the endogenous selection of heterogeneous firms in an industry” and that his model set up “…embeds firm productivity heterogeneity within Krugman’s model of trade under monopolistic competition and increasing returns” (see Melitz 2003, p. 1696).

3.1 Autarky

Melitz (2003) uses the same general equilibrium framework as in Krugman (1980). In Melitz, consumers are characterized as having CES preferences which reflect their love of (differentiated) variety for goods as in Krugman (1980) (this elasticity of substitution between varieties is given by σ). Welfare depends on the number of varieties available to consumers and on the average price level in the economy. The major difference from Krugman (1980) is on the production side. There are a large number of firms in the market, each producing a horizontally differentiated good under increasing returns using labour as the only factor of production. Heterogeneity is modelled by differences among firms in their (labour) productivity (or the marginal product of labour), denoted φ. Obviously, the higher the productivity of the firm the lower the marginal costs of production.[16] All firms share the same fixed costs of production denoted by f, but because of differences in labour productivity, have different marginal costs of production. In this economy, there are an unbounded number of potential firms at any moment who want to enter the market. However, entry is not free and firms have to pay a sunk entry cost, denoted by f_e, measured in terms of labour, to enter the market.[17] Firms do not know their productivity before entering the market. The productivity levels are assumed to be drawn by firms from a distribution characterized by a cdf $G(\varphi)$ (pdf $g(\varphi)$). Firms can draw ("discover") their productivity only after paying the sunk cost f_e. So, low- and high-productivity firms can coexist in the market. In this setup, the ratios of any two firms' outputs and revenues depend directly only on the ratio of their productivity levels (and the price ratios depend inversely on the ratio of their productivity levels).[18] Therefore, in line with the empirical evidence, more productive firms sell more output, generate higher revenues and make more profits than less productive firms.[19]

Active firms may exit the market if they are hit with an exogenous shock (like a sudden fall in demand) with probability δ.[20] After a firm draws its productivity from the productivity distribution, it makes a decision of whether it wants to produce or whether it wants to exit the market depending on whether its expected value of the discounted sum of profits (discounted by δ) are high enough to repay the initial fixed entry cost f_e. A stationary equilibrium for this economy exists when the

---

[16]So marginal costs are $\frac{1}{\varphi}$.

[17]The sunk cost f_e could be the cost of carrying out a market survey or the development cost of a new variety of a good before the firms actually enter the market.

[18]See Melitz (2003, Eqs. 3 and 6, pp. 1699–1700).

[19]All of these variables ultimately depend on the firm's productivity. Formally, profits (π) and revenue (r) depend directly on the firm's productivity and are denoted $\pi(\varphi)$ and $r(\varphi)$ respectively; prices depend inversely on a firm's productivity and is denoted as $p(\varphi)$. $\pi(\varphi)$ and $r(\varphi)$ also depend on aggregate price and revenue in the economy (defined later).

[20]This probability is a constant across all firms and over all periods and is modelled after Hopenhayn (1992a, b).

number of incumbent firms that are forced to exit in this manner is equal to the number of firms that are able to successfully enter the market. Melitz defines a "cut-off" productivity level φ^*, such that the (marginal) firm with this productivity level just makes zero expected profits, or $\pi(\varphi^*) = 0$ and any firm with $\varphi < \varphi^*$ exits the market. So the productivity cut-off φ^* is the minimum productivity that ensures a positive value of profit for the firm. Only firms that draw a productivity $\varphi > \varphi^*$ decide to produce. This shapes the ex-post probability distribution of productivity for the firms. Since only firms with productivity $\varphi > \varphi^*$ stay in the market, the actual (ex-post) pdf distribution $\mu(\varphi)$ is a truncated distribution given by $\frac{g(\varphi)}{1 - G(\varphi^*)}$. Therefore, the ex-post pdf of productivity for firms that survive to produce for the market is derived from the *ex-ante* pdf of entering firms via the entry–exit mechanism mentioned earlier.

Now Melitz introduces a number of aggregate variables to make the analysis tractable in this heterogeneous productivity setup. If there are *M* firms that actually produce in the market, then given this mass of firms, it is easy to determine the values for aggregate variables in this economy like the overall price level *P*, total profits Π, and total production level for all differentiated goods *Q* (see Melitz 2003, p. 1700). Now, instead of dealing with this mass of heterogenous firms, one can conveniently think of a "representative firm" in this setup. This representative firm is a firm with a productivity level equal to the weighted average productivity level of all the surviving firms' productivity levels (calculated using the aforementioned ex-post distribution of productivity $\mu(\varphi)$). Melitz denotes this average productivity level as $\tilde{\varphi}$. The representative firm with productivity level $\tilde{\varphi}$ earns "average" profits denoted by $\bar{\pi}$, such that the average profit times the number of active firms in this economy gives the total profits in this economy $\left(\text{or } \bar{\pi} = \frac{\Pi}{M}\right)$.[21] Similarly, this representative firm will induce (on multiplying by *M*) the same aggregate price level, revenue and quantity as the economy with *M* heterogeneous firms with the ex-post productivity distribution $\mu(\varphi)$. Since the ex-post productivity distribution, $\mu(\varphi)$, itself depends on the cut-off productivity level φ^*, the measure of the (ex-post) average productivity $\tilde{\varphi}$ mentioned above can now be obtained as a function of this cut-off φ^* denoted by $\tilde{\varphi}(\varphi^*)$. Using the ex-post productivity distribution, it can be shown that average profit, average revenues and the average price level are also ultimately determined by the cut-off productivity level (see Melitz 2003, p. 1703). Of these variables, the most important variable that is subsequently used is the average profit $\bar{\pi}$, which can be denoted as a function of the average productivity $\tilde{\varphi}(\varphi^*)$ or as $\bar{\pi}[\tilde{\varphi}(\varphi^*)]$.

Two key relations are then used to endogenously determine the actual values of the cut-off productivity level φ^* and the average profit $\bar{\pi}$ in the economy. The first relation is the zero-profit cut-off condition (or ZCP) and the second is the free entry (FE) condition. The ZCP condition gives a negative relation between $\bar{\pi}$ and φ^*.

[21] Note that since all firms that are active in the market earn positive profits (except the marginal firm that earns zero profits) the average level of profits $\bar{\pi}$ must be positive. Positive profits attract firms to the industry.

When the cut-off productivity φ^* increases, two opposing effects come into play. Each surviving firm is now more productive since the cut-off survival productivity level is higher. Since average profit is a function of cut-off productivity level, average profit has a tendency to increase. But by the same reasoning, all the other surviving firms that have productivity levels higher than the now increased cut-off productivity level are also now more productive. Therefore there is more competition among firms for profits, which causes average profits to decrease. Under some mild assumptions on the probability distribution $G(\varphi)$, the second effect dominates the first effect giving a net negative relation.[22] The FE condition gives a positive relation between $\bar{\pi}$ and φ^*. The FE condition states that in equilibrium the expected value of the future stream of profits for a firm should be equal to the fixed cost of entry so that the net value from entering the market is zero. As the cut-off productivity increases, fewer firms will be able to enter the market. The entering firms will be higher productivity firms that will earn higher profits leading to higher average profits (conditional on entry).

In the (π, φ) space, the positively sloped FE and the negatively sloped ZCP curves intersect once giving the equilibrium average profit $\bar{\pi}$ and the equilibrium cut-off productivity φ^* in autarky. To close the model, Melitz uses a labour market clearing equation. In Melitz as in Krugman (1980), labour is the only factor of production which is used either for production or to pay the fixed costs of entry. As labour is the only factor of production, all income accrues to labour and therefore total income equals total expenditure on differentiated goods by labour-consumers. Using the equilibrium values of $\bar{\pi}$ and φ^* derived earlier and the labour market clearing condition, Melitz determines the number of varieties/firms, total output, aggregate prices and welfare. With wages of labour normalized to 1, welfare is inversely related to the aggregate price level.

3.2 Trade

Now Melitz considers the effects of trade in this economy. He assumes that the economy under study begins to trade with $n \geq 1$ other countries. If trade costs are absent, then under trade, equilibrium would just mean a proportional increase in the scale of this economy without any effect on a firm's revenue, profits, etc. However, Melitz introduces iceberg transport costs *a la* Samuelson (1952) denoted $\tau > 1$ which is the amount of any variety that a firm has to produce in order to ship 1 unit of a variety to a destination. More importantly, he also adds a fixed cost of exporting, denoted as f_x, which is a lump-sum cost to start exporting and which is

[22]Formally, as noted in Melitz (2003, p. 1704), a sufficient condition for the ZCP to decrease is that $\frac{g(\varphi)\varphi}{1 - G(\varphi)}$ should be increasing to infinity on $(0, \infty)$—an assumption which holds for many probability distributions.

paid only after the firm learns about its productivity.[23] Since exporting is costly there is a further selection of better-performing firms into the export market. Further, since firms can decide to export only after drawing their productivity from $g(\varphi)$, all exporters sell their goods in the domestic market but not all firms in the domestic market can export. Melitz defines another productivity cut-off, denoted as φ_x^*, such that firms with productivity levels above φ_x^* earn profits from selling in the domestic market and these firms can also earn profits from exporting the goods to the foreign market.[24] Firms with $\varphi^* < \varphi < \varphi_x^*$ produce exclusively for the domestic market. Melitz shows that this partitioning of firms—into firms producing exclusively for the domestic market, and firms producing for both domestic and foreign markets—is only possible under the condition that the fixed costs of exporting (f_x) are sufficiently high.[25] Since $\varphi_x^* > \varphi^*$ only the most productive firms export. Under trade, Melitz recomputes the ZCP and FE conditions for the integrated market. The FE condition remains unchanged but the ZCP curve shifts upwards since the firms that survive on average make higher profits. Therefore, the cut-off productivity level increases under trade.

Melitz now considers the implications of trade on the allocation of market shares and profits among firms and on aggregate productivity. Comparing autarky with free trade, he shows that post-trade the least-productive firms leave the domestic market (a domestic market selection effect) while high-productivity firms enter the export market (an export market selection effect) and these two kinds of selection effects work to drive market shares towards more productive firms. The mechanism through which these effects operate is the following. Under trade, all firms suffer a loss in domestic sales (Melitz 2003, p. 1714). This causes the least-productive firms to exit the market as they are unable to earn positive profits. However, exporting firms make up for the loss in domestic sales with foreign sales. These firms increase their production to exploit the opportunity of earning additional profits from exports and hence increase their demand for labour. Demand for labour from these new exporters increases the overall demand for labour in the economy and causes a rise in the real wage. As a result of this increased wage, some of the less-productive firms that were just breaking even, now make losses and are forced to exit the market.[26] The net result of these selection forces is a reallocation of market shares from low-productivity firms to high-productivity firms leading to an increase in the average productivity of this economy. Melitz recognizes that this mechanism "… highlights a potentially important channel for the redistributive effects of trade within industries" (Melitz 2003, p. 1716). He notes that such a restructuring of the

---

[23]Examples of such costs include fixed costs like the cost of setting up a distribution channel in a foreign market, or informing foreign consumers about a product, or a production cost incurred to modify home products for foreign consumption or to confirm to foreign regulatory standards.

[24]Exporters are also more productive and bigger as a consequence of the higher productivity. This is in line with empirical evidence.

[25]Formally, this condition is $\tau^{\sigma-1} f_x > f$. A large fixed cost of exporting will give $f_x > f$ and lead to this partitioning.

[26]See Sect. 7.2 in Melitz (2003, p. 1715).

economy might have contributed to the increase in productivity in U.S. manufacturing reported in Bernard and Jensen (2004). So, the model is also successful in explaining the increase in post-liberalization productivity observed in the data. In sum, therefore, the Melitz model explains quite a few stylized facts observed in the data that could not be explained by new trade theory.

4 Subsequent Research

The Melitz model has initiated a large literature which explores the implications of incorporating firm heterogeneity in a number of different setups. In the paragraphs below, we discuss some of the main extensions/applications of the Melitz framework. Since the literature following Melitz is quite extensive, we select only the direct and most important extensions/applications of Melitz's framework. We consider in the paragraphs below three important extensions: (i) the extension of Melitz to cover the case of internationalization of the firms via FDI (Helpman et al. 2004) (ii) the extension of the Melitz framework to consider the effect on trade flows following the removal of trade barriers (Chaney 2008), and (iii) the extension of the Melitz framework to explain certain features of the bilateral trade matrix, thereby leading to a more precise gravity equation (Helpman et al. 2008). We also mention some of the other related literature and a model that has been developed as an "alternative" to the Melitz model.

We start with Helpman et al. (2004) who extend the Melitz framework to include FDI and trade. They answer the question of the mode of foreign market access in the Melitz setup, that is, why some firms become exporters and why other firms choose to serve foreign markets via FDI. In their model, in any sector there is a cost of market entry f_E, a cost of production in the domestic market f_D, a fixed cost of exporting f_X (for each foreign market), and there is also an additional fixed cost of FDI given by f_I. Unlike in Melitz, the marginal costs of production are the same for all firms. Their critical assumption is that f_I adjusted for relative wages w_i and w_j in two countries i and j) is greater than f_X adjusted for trade costs τ_{ij} between countries i and j which in turn is greater than f_D (see Helpman et al. 2004, Eq. (1) on p. 302). This condition ensures that firms sort themselves into a hierarchy—the least-productive firms exit the market altogether, low-productivity firms produce only for the domestic market, more productive firms absorb transport costs and export to the foreign market and the most productive firms incur the fixed FDI costs to set up a subsidiary in another country and become MNEs.[27] The authors compute the productivity cut-offs at which these entry mode switches occur and then they take their model to data.[28] Their empirical analysis is conducted at the sectoral level

[27]Using the COMPUSTAT database for the US for 1996, Helpman et al. (2004) show (Table 1 on p. 301) that MNE's are 15 % more productive than even exporters.

[28]Helpman et al. (2004, p. 303, Eqs. (2)–(4)).

using data for 52 different manufacturing sectors for the U.S. for 1994. Their main variable of interest is denoted by $\frac{s_X^{ij}}{s_I^{ij}}$ or the ratio of exports (X) to local FDI sales (I) for a pair of countries i (the U.S.) and j. They regress this variable on transport costs (denoted by the variables FREIGHT and TARIFF in their paper which represent ad-valorem measures of freight and insurance costs and trade taxes), a measure of plant-level fixed costs (denoted FP in their paper), country dummies to capture country-specific fixed costs (same for both export sales and FDI), sectoral capital, and R&D intensities (denoted KL and RD, respectively) and most importantly, on the degree of intra-industry firm heterogeneity which is captured by the dispersion of firm size within a sector (denoted DISPERSE). This last variable used in their study requires some explanation. Helpman et al. (2004) assume that firms' productivity is drawn from a Pareto distribution.[29] The Pareto distribution has a number of tractable analytical properties one of which is the fact that under this distribution an observed dispersion measure of domestic sales of firms can be used as a proxy for the underlying productivity distribution of firms.[30] Given this property of the distribution, the authors predict that an increase in the observed dispersion measure in a sector (i.e. an increase in productivity heterogeneity in that sector) should lead to relatively more FDI than exports in that sector or a decrease in their LHS variable. Their empirical results indicate that the coefficient of dispersion (or DISPERSE) is significantly negative which indicates that productivity heterogeneity reduces exports relative to FDI as expected.[31]

The second major extension of the Melitz framework is by Chaney (2008) who derives implications of the Melitz model for the elasticity of trade flows to trade barriers. Chaney uses a simplified Melitz-type heterogenous firm setup and comes up with a number of results regarding this elasticity that are very different from the results obtained under new trade theory which assumes identical firms (Krugman 1980). First, he shows that in response to a lowering of trade barriers not only do existing exporters export more (an increase in the intensive margin) but the set of exporters increases as well (an increase in the extensive margin). This result is different from Krugman (1980) where only the intensive margin of trade is affected following a reduction in trade barriers. The additional margin of adjustment in Chaney (2008) means that the elasticity of trade flows to trade barriers is much larger in magnitude than in Krugman (1980). According to Chaney, given the observed distribution of firm sizes in the U.S., this elasticity is likely to be twice as

[29]See Helpman et al. (2004, p. 304) for details. The cumulative distribution function for a Pareto distribution for a random variable X (with scale parameter b and shape parameter k) is given by $F(x) = 1 - (\frac{b}{x})^k, x \geq b > 0$.

[30]Formally, if the underlying productivity is drawn from the Pareto distribution, then the distribution of firm domestic sales is $k - (\varepsilon - 1)$ where ε is the elasticity of substitution between goods in a CES utility function.

[31]They also find that the LHS ratio depends negatively on the variables FREIGHT and TARIFF, positively on FP and negatively on KL and RD as expected.

large as previously predicted. Chaney also derives an important result relating the elasticity of trade flows with respect to trade barriers and the elasticity of substitution in the CES demand function. Following Chaney, let the total exports from country i to j be denoted by X_{ij}. Firms draw their productivity from a Pareto distribution with scale parameter γ. There are two costs associated with exporting. First is the standard variable iceberg transport cost of exporting τ_{ij} and the second is the fixed cost of exporting f_{ij} (for countries i and j). Consumer preferences are CES with elasticity of substitution given by σ.[32] Chaney defines two types of trade flow elasticities—the first is the elasticity of trade flows with respect to the variable costs of exporting (or variable trade barriers) denoted by $\zeta \equiv -\frac{\mathrm{d}\ln\ X_{ij}}{\mathrm{d}\ln \tau_{ij}}$ and the second is the elasticity of trade flows with respect to the fixed costs of exporting (or fixed trade barriers) denoted $\xi \equiv -\frac{\mathrm{d}\ln X_{ij}}{\mathrm{d}\ln f_{ij}}$. Each of these elasticities has an intensive margin term and an extensive margin term also expressed as elasticities (for details see Chaney (2008, pp. 1716–1717). Chaney is interested in the impact of the elasticity of substitution σ on these two elasticities or $\frac{\partial\zeta}{\partial\sigma}$ and $\frac{\partial\xi}{\partial\sigma}$. He finds that for a specific distribution used in the analysis (Pareto), $\frac{\partial\zeta}{\partial\sigma} = 0$ and $\frac{\partial\xi}{\partial\sigma} < 0$. To see why this is so let us first consider $\frac{\partial\zeta}{\partial\sigma}$. Recall that ζ is the elasticity of trade flows with respect to variable trade costs τ_{ij}. Chaney shows that when τ_{ij} changes both the intensive as well as the extensive margins change. Formally, the intensive margin elasticity is given by $(1 - \sigma)$ which is the same as in Krugman (1980). However, as mentioned earlier, in addition to the change in the intensive margin there is also the extensive margin of adjustment actuated by a change in τ_{ij}. This elasticity is $\gamma - (\sigma - 1)$. In sum the intensive and the extensive margin elasticities add up to γ which implies that on aggregate the overall elasticity ζ is independent of σ giving $\frac{\partial\zeta}{\partial\sigma} = 0$. The intuition behind this result is the following. Industries which have a high elasticity of substitution or a high σ are industries that are characterized by a high level of competition.[33] In these industries, less-productive firms have small market shares and high-productivity firms have large market shares. When variable trade barriers fall, incumbent high productivity exporters increase their market share by exporting more to foreign markets (affecting the intensive margin) while the low-productivity firms export very little (so that there is little or no effect on the extensive margin). In contrast, in industries that have a low σ (less competition), low-productivity firms manage to have larger market shares. In these industries, when low-productivity firms start exporting there is a considerable change in the extensive margin but the intensive margin does not change much. In either case (high or low σ) one of the margins is more responsive and the other margin is less responsive giving a resultant elasticity of 0. A similar intuition holds for the case of

[32] Note that $\gamma > \sigma - 1$ (Chaney 2008, p. 1711, footnote 6).

[33] A high σ means that the differentiated varieties of a good are more substitutable.

the elasticity of trade flows with respect to fixed trade costs.[34] In general, in a heterogeneous firm setup, the elasticity of trade flows with respect to trade barriers will be high in those sectors which have a low elasticity of substitution. Chaney's formulation lends itself quite easily to an econometric investigation. In an earlier version of the paper (Chaney 2005), he makes use of data for 169 countries and 265 sectors for the period 1980–1997 on bilateral trade flows, different measures of trade barriers/costs (distance, contiguity, etc.) and estimates of sectoral-level elasticities of substitution (computed in an earlier draft of the paper by Broda and Weinstein (2006)). He finds strong empirical support for his model.

Finally, we discuss Helpman et al. (2008) who derive implications of the Melitz framework for the extensive margin of trade as given by the gravity equation. Their model is motivated by the search for an explanation of the large number of zeroes observed in the bilateral trade matrix (zero trade between country pairs). They argue that the standard methods of estimating the gravity equation which ignore the zeroes in trade data by discarding them from the estimation sample (by treating them as missing observations) or which impose symmetry restrictions are flawed and give biased estimates. They build a generalized version of the Melitz model similar to Chaney (2008). The setup is the same as in Melitz (2003), where firms face fixed and variable costs of exporting. As in Melitz, only the more productive firms can break into the export market. However, unlike in Melitz where the mass of entrants is unbounded, Helpman et al. (2008) impose bounds on the distribution of productivity of firms. Another innovation in this framework is that the profitability of exporting can vary over destination countries (because of differences in fixed and/or variable export costs and demand conditions across countries). In this framework, for two countries i and j there is a lower productivity bound for firms in country i at which a firm can just break even by exporting to j. A firm in i has to draw a productivity greater than this bound to earn a positive profit from exporting to j. It is possible that there is no firm in i with a productivity over this bound leading to zero trade flows between i and j. Moreover, it is possible that a firm in i has a productivity level over the cut-off productivity for another destination j' leading to positive trade flows between i and j'.[35] So this model predicts positive as well as zero trade flows across pairs of countries and allows for the number of exporting firms to vary across destination countries. Helpman et al. (2008) then use a two-stage estimation process to estimate a "precise" gravity equation where they correct for the self-selection of firms into export markets and for potential asymmetries in trade flows between pairs of countries.[36] They can estimate the effect of

[34]There are some differences with the variable trade costs case. For a change in fixed trade costs, the intensive margin elasticity is 0, since a change in the fixed costs does not affect the export decisions of incumbent exporters. The extensive margin elasticity is $\frac{\gamma}{\sigma-1}-1$ giving $\frac{\partial \xi}{\partial \sigma}<0$.

[35]See Helpman et al. (2008, pp. 441–444, 449–452).

[36]They use observed data from the bilateral trade matrix for 158 countries in 1986 to infer the magnitude of the trade costs between countries. This information allows them to estimate the gravity equation in the second stage.

trade frictions on the intensive (per firm exports) and extensive (number of exporters) margins of trade. Their empirical results show that the traditional estimates of gravity equation are biased which happens mainly due to the omission of the extensive margin.

Apart from these major extensions the Melitz heterogeneous firm framework has been used to examine interesting questions in other domains. Helpman et al. (2010) extend the Melitz model to an analysis of the labour market. Arkolakis and Muendler (2010) use the Melitz model to examine the role of marketing costs in a heterogeneous firm set up. Chaney (2013), Muûls and Pisu (2009), Manova (2008, 2013) (among others) use/extend the Melitz setup to study the effect of credit constraints in a heterogeneous firm setting.

With a large number of applications in almost every field in international economics, the Melitz model has now become the benchmark model in the new-new trade theory framework. Before concluding, we mention a literature which forms an "alternative" to the Melitz framework. This is the work by Bernard et al. (2003) (BEJK hereafter). BEJK develop an alternative model incorporating firm heterogeneity that also explains many of the stylized facts mentioned earlier. The work by BEJK builds on the framework of an earlier paper by Eaton and Kortum (2002) (EK hereafter). We first briefly discuss the model in EK and then discuss the paper by BEJK.

EK are motivated by the following stylized facts: (i) trade diminishes dramatically with distance (ii) prices vary across locations with greater differences between places farther apart (iii) factor rewards vary across countries (iv) countries' relative productivities vary substantially across industries. EK provide a unified model to explain the above stylized facts. They construct a multi-country, multi-industry Ricardian model with geographical barriers which not only explains the above-mentioned stylized facts but also is useful in understanding issues such as the gains from trade, the role of trade in spreading the benefits of technological progress and the implications of tariff liberalization.[37] The Ricardian element of comparative advantage arising from technological differences promotes trade while geographical barriers inhibit trade. The key modelling innovation of EK is to model technology probabilistically which allows them to extend the continuum Ricardian model to a multi-country setting in a tractable way. The productivity of a country i in producing goods ω is a random variable z_i which is drawn (independently for each of goods ω) from a Frechet distribution with parameter $\theta > 1$ (same for all countries), or formally from $F_i(z) = Pr(Z_i \leq z) = e^{-T_i z^{-\theta}}$. In a trade context, T_i can be thought of as reflecting the technology of country i and reflects absolute advantage while θ determines the heterogeneity in the productivity across different goods and thus denotes comparative advantage in a probabilistic sense (see Eaton and Kortum

[37]EK extend the work by earlier Dornbusch et al. (1977). The latter work is itself an extension of the Ricardian model to a continuum of goods but only for two countries. EK generalize this work by extending the Ricardian general equilibrium framework with a continuum of goods to many countries and they also allow a role for economic geography to influence trade flows.

2002, p. 1747 for details). Geographical barriers are introduced in the form an iceberg cost, d_{ni}, between countries n and i ($d_{ni} > 1$ for $i \neq n, d_{ii} = 1$ for all i). An implication of the probabilistic technology assumption combined with a continuum of goods is that the fraction of goods that country n buys from country i is also the fraction of its expenditure on goods from country i. This, in turn, delivers a gravity model type prediction that bilateral trade volumes depend positively on the GDP of the two countries and negatively on distance. A deeper contribution of the model is to go beyond the standard gravity variables and uncover structural parameters governing the roles of technology and geographical barriers. The parameter estimates allow them to perform a number of counterfactual exercises: all countries benefit from free trade; the gains from moving to a world with no geographic barriers are enormous; an improvement in a country's technology raises welfare in all countries with gains being greater in countries enjoying proximity to the source; all countries benefit from a multi-lateral move to free trade. To sum up, not only does the EK paper provide an elegant model to incorporate the stylized facts mentioned above, it is the first paper which models heterogenous sectors in a general equilibrium setup. In doing so it lays the groundwork for subsequent work (e.g. Melitz 2003; Bernard et al. 2003) doing a quantitative analysis of trade when heterogeneity is important.

The work by Bernard et al. (2003) (BEJK hereafter) builds on the framework in EK. A key departure from EK is to explicitly introduce firms into the analysis and assume that firms are engaged in a Bertrand competition as opposed to the perfectly competitive markets in EK. Since preferences are CES giving rise to a constant elasticity demand function, the successful seller (the most efficient one) in any market charges a price which equals the minimum of the unit cost of the second most efficient producer and the mark-up price of the most efficient producer. Since the price charged for each goods in each market depends on the unit costs of the two most efficient producers for each goods, BEJK assume that the productivity of top two producers for any goods ω in country i is a pair of random variables $z_{1i}(\omega)$, $z_{2i}(\omega)$ with a joint distribution analogous to Frechet distribution in EK.[38] An important implication is that the mark-up has a distribution which is same across destinations. As well, low-cost producers are more likely to charge a higher mark-up, and their measured productivity is likely to be higher. Generating variable mark-ups across producers resulting in different measured productivity was one of the key motivations behind departing from the competitive market structure in EK. Additionally, given the iceberg trading cost, it is more difficult to export than to sell domestically. An implication is that any exporting firm must sell domestically as well, but not all domestically active firms will succeed in exporting. This is consistent with the stylized fact mentioned earlier that only a small fraction of active firms engage in exporting. The paper also develops a simulation approach to evaluate how well the model does quantitatively in explaining the stylized facts.

[38]Formally, the productivity of top two producers in country-i, Z_{1i} and Z_{2i} have the joint distribution: $\Pr(Z_{1i} \leq z_1, Z_{2i} \leq z_2) = F_i(z_1, z_2) = [1 + T_i(z_2^{-\theta} - z_1^{-\theta})]e^{-T_i z_2^{-\theta}}$.

Comparing BEJK to Melitz, note that the demand side is the same in the two characterized by CES preferences. The difference comes in the market organization. Unlike the monopolistic competitive framework in Melitz, in BEJK firms engage in Bertrand price competition which results in variable mark-ups of price over marginal costs despite CES preferences. This has implications for welfare discussed below.

5 Welfare Implications

We now look at the welfare implications of trade liberalization in some of the models mentioned above.[39] In Melitz, the welfare gains from trade liberalization come from an increase in aggregate productivity due to the reallocation of resources in favour of more productive firms. However, there is a mechanism through which welfare may also decrease. This works through the number of product varieties consumed under trade. Upon opening to trade, although some domestic firms are forced to exit the market, consumers usually enjoy more varieties in total because the number of new foreign varieties that they have access to typically is greater than the number of domestic varieties that they are forced to forsake. However, the possibility of the number of varieties available to domestic consumers declining, which would be a source of welfare loss, exists. Melitz shows that the welfare enhancing effects of the increase in aggregate productivity dominates the welfare reducing effects (if any) of the loss of variety. Melitz (2003) is the first paper to point out the welfare gains stemming from the increase in aggregate productivity due to this reallocation of market shares towards high-productivity firms. Similar results were obtained by BEJK (2003) (BEJK). However, there are some additional sources of gains from trade in BEJK. One of them arises due to the endogeneity of mark-ups. Increased product market competition induced by trade liberalization reduces mark-ups which is an additional source of gain from trade.[40] Another source of gains from trade involves the use of intermediate inputs in the production process. In BEJK the relative price of these intermediate inputs decreases under trade as cheaper imported inputs replace domestically produced inputs leading to an additional pathway through which productivity and welfare can increase.

All of the different channels (mentioned above) through which welfare gains accrue naturally lead to the question of what are the actual aggregate gains from trade. Arkolakis et al. (2008, 2012) (ACR) attempt to answer this question. Arkolakis et al. (2008) show that the major quantitative frameworks in international trade deliver comparable expressions for welfare. Arkolakis et al. (2012) then show

[39]For details and for a fuller exposition see Bekkers (2008, pp. 16–33), which gives an excellent comparison of the welfare implications of several models in the Melitz tradition.

[40]Melitz and Ottaviano (2008) develop an alternative framework with variable mark-ups by using a linear demand function. Similar to BEJK trade liberalization provides additional gains through a reduction in mark-ups.

that for a broad class of models it is possible to answer the question of aggregate welfare gains using readily available data. They show that the change in welfare (or real income), $\widehat{W}_j$ for country j, from (say) a trade liberalization episode can be written as a function of two variables (sufficient statistics): (1) the share of expenditure on domestic goods λ_{jj} and (2) the elasticity of relative imports with respect to variable trade costs ε. Formally, $\widehat{W}_j = \hat{\lambda}_{jj}^{\frac{1}{\varepsilon}}$,where$\widehat{\ }$denotes the change in a variable value between two equilibria in response to the "shock" of trade liberalization (for details see Arkolakis et al. 2012, p. 99). Using this expression and estimates of ε from gravity models of trade they estimate the gains from trade for the U.S. for the year 2000 at 0.7–1.4 %. Although their expression for welfare is derived under a number of simplifying assumptions (see the section on "Macro-Level Restrictions", Arkolakis et al. 2012, pp. 101–104), the welfare gains are very low and quite puzzling in light of the additional channels of welfare gains emerging in the Melitz model and later models. In a response to ACR, Melitz and Redding (2014) compare heterogeneous and homogenous firm models (holding all other structural parameters constant) where heterogeneous firms make the endogenous decision of whether to enter or to leave the export and domestic markets. This entry and exit decision leads to endogenous changes in aggregate productivity and to positive welfare gains absent in models without firm heterogeneity. They find on calibrating their model to U.S. firm-level data that this channel contributes significantly to welfare. They point out that in ACR the restrictions imposed on the heterogenous firm models lead to the counter-intuitive results noted above. However, under small and plausible deviations from these restrictions, the underlying microstructure does have an effect on welfare calculations and λ and ε above are no longer sufficient statistics from which one can calculate welfare.

6 Conclusion

The class of new-new trade theory models gives a central role to the firm in mediating trade flows. These models explain the stylized facts observed in firm micro-data. These models have bridged the gap between theoretical models of trade flows and the observed data considerably. Moreover, these models of trade are quite tractable and amenable to modification in a variety of settings. They have been and are being used to model and answer interesting questions in a number of settings in various fields of economics.

References

Arkolakis, C, Muendler M-A (2010) The extensive margin of exporting products: a firm-level analysis. Discussion paper, National Bureau of Economic Research

Arkolakis C, Demidova S, Klenow PJ, Rodriguez-Clare A (2008) Endogenous variety and the gains from trade. American Econ Rev Papers Proc 98(4):444–450

Arkolakis C, Constinot A, Rodriguez-Clare A (2012) New trade models, same old gains? Am Econ Rev 102(1):94–130
Aw B, Hwang A (1995) Productivity and the export market: a firm-level analysis. J Dev Econ 47:313–332
Aw B, Chung S, Roberts M (2000) Productivity and turnover in the export market:micro-level evidence from Republic of Korea and Taiwan (China). World Bank Econ Rev 14(1):65–90
Bekkers E (2008) Essays on firm heterogeneity and quality in international trade, no. 427. Tinbergen Institute Research Series, Tinbergen Institute
Bernard A, Jensen J (1995) Exporters, jobs and wages in U.S. manufacturing, 1976–1987. Brookings papers on economic activity, Microeconomics
Bernard A, Jensen JB (1999) Exceptional exporter performance: cause, effect or both? J Int Econ 47(1):1–25
Bernard A, Jensen JB (2004) Exporting and productivity in the USA. Oxford Rev Econ Policy 20(3):343–357
Bernard AJ, Eaton JB, Jensen S, Kortum S (2003) Plants and productivity in international trade. Am Econ Rev 93(4):1268–1290
Bernard AB, Jensen JB, Redding SJ, Schott PK (2007) Firms in international trade. J Econ Perspect 21:105–130
Bowen HP, Leamer EE, Sveikauskas L (1987) Multicountry, multifactor tests of the factor abundance theory. American Econ Rev 791–809
Broda C, Weinstein D (2006) Globalization and the gains from variety. Q J Econ 121(2):541–585
Chaney T (2005) Distorted gravity: the intensive and extensive margins of international trade. MIT, Department of Economics
Chaney T (2008) Distorted gravity: the intensive and extensive margins of international trade. Am Econ Rev 98(4):1707–1721
Chaney T (2013) Liquidity constrained exporters. Discussion paper, National Bureau of Economic Research
Clerides S, Lach S, Tybout J (1998) Is learning by exporting important? Microdynamic evidence from Columbia, Morocco. Quart J Econ 113(3):903–947
Dornbusch R, Fischer S, Samuelson P (1977) Comparative advantage, trade, and payments in a ricardian model with a continuum of goods. Am Econ Rev 67(5):823–839
Eaton J, Kortum S (2002) Technology, geography and trade. Econometrica 70:1741–1779
Grubel HG, Lloyd P (1975) Intra-industry trade: the theory and measurement of international trade in differentiated products. MacMillan, London
Helpman E, Krugman P (1985) Market structure and foreign trade: increasing returns, imperfect competition, and the international economy. The MIT Press
Helpman E, Melitz MJ, Yeaple SR (2004) Export versus FDI with heterogeneous firms. Am Econ Rev 94(1):300–316
Helpman E, Melitz M, Rubinstein Y (2008) Estimating trade volumes: trading partners and trading volumes. Quart J Econ 2:441–487
Helpman E, Itskhoki O, Redding S (2010) Inequality and unemployment in a global economy. Econometrica 78(4):1239–1283
Hopenhayn H (1992a) Entry, exit and firm dynamics in long run equilibrium. 60:1127–1150
Hopenhayn H (1992b) Exit, selection, and the value of firms. 16:621–653
Krugman P (1979) Increasing returns, monopolistic competition and international trade. J Int Econ 9:469–479
Krugman P (1980) Scale economies, product differentiation, and the pattern of trade. Am Econ Rev 70(5):950–959
Krugman P (2008) The increasing returns revolution in trade and geography. Nobel Prize Lecture, pp 335–348
Manova K (2008) Credit constraints, equity market liberalizations and international trade. J Int Econ 76(1):33–47
Manova K (2013) Credit constraints, heterogeneous firms, and international trade. Rev Econ Stud 80(2):711–744

Mayer T, Ottaviano G (2008) The happy few: the internationalisation of european firms-new facts based on firm level evidence. Intereconomics 43(3):135–148

Melitz MJ (2003) The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71(6):1695–1725

Melitz MJ, Ottaviano GI (2008) Market size, trade, and productivity. Rev Econ Stud 75(1):295–316

Melitz MJ, Redding SJ (2014) New trade models, new welfare implications. NBER working paper

Muûls M, Pisu M (2009) Imports and exports at the level of the firm: evidence from Belgium. World Econ 32(5):692–734

Olley G, Pakes A (1996) The dynamics of productivity in the telecommunications equipment industry. Econometrica 64(6):1263–1297

Pavcnik N (2002) Trade liberalization, exit, and productivity improvement: evidence from chilean plants. Rev Econ Stud 69(1):245–276

Ranjan P, Raychaudhuri J (2011) Self-selection vs learning: evidence from Indian exporting firms. Indian Growth Dev Rev 4(1):22–37

Samuelson P (1952) The transfer problem and transport costs: the terms of trade when impediments are absent. Econ J 62:278–304

Trefler D (1995) The case of the missing trade and other mysteries. American Econ Rev 1029–1046

Trefler D (2004) The long and short of the Canada-U.S. free trade agreement. Am Econ Rev 94(4):870–895

Van Biesebroeck J (2005) Exporting raises productivity in sub-saharan african manufacturing firms. J Int Econ 67(2):373–391

Time Zones and FDI with Heterogeneous Firms

Toru Kikuchi, Sugata Marjit and Biswajit Mandal

Abstract Based on Helpman et al. (Am EconRev 94:300–316, 2004), we propose a simple two-country (Home and Foreign) model with heterogeneous firms to capture the role of FDI via utilizing time zone differences. Two countries are located in different time zones and there is no overlap in daily working hours. It will be shown that productivities of the firms undertaking FDI are higher than the productivities of non-FDI firms. Although the results look quite similar with Helpman et al. (Am EconRev 94:300–316, 2004), the direction of service trade flow is totally different: foreign subsidiaries of high-productivity firms serve the Home market.

Keywords Time zones · FDI · Heterogeneous firms

JEL classification F12

In this paper, we are essentially completing the work for Late Toru Kikuchi who was instrumental in developing the basic idea. Sugata Marjit acknowledges the financial assistance from the RBI endowment at CSSSC. Biswajit Mandal thankfully acknowledges the hospitality provided by the University at Albany-SUNY during his visit as C.V. Raman fellow. We also express thanks to the referee for some helpful comments and suggestions. However, we retain sole responsibility for any remaining errors.

T. Kikuchi (Deceased)
Graduate School of Economics, Kobe University, Kobe, Japan

S. Marjit (✉)
Centre for Studies in Social Sciences, Kolkata, India
e-mail: marjit@gmail.com

S. Marjit
Calcutta University, Kolkata, India

B. Mandal (✉)
Department of Economics and Politics, Visva-Bharati University, Santiniketan, India
e-mail: biswajiteco@gmail.com

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_2

1 Introduction

Since 1980s, foreign direct investment (FDI) has grown astonishingly fast, even faster than international trade. Not only did the overall level of FDI increase, it has also been changed from investments in manufacturing to investment in services. Related to these, intra-firm trade of business services such as engineering, consulting, and software development that do not require physical shipments of products, have been playing major roles.[1]

Following these changes, new types of FDI and service trade surfaced in the recent past. Such investment and trade are taking advantage of time zone differences between countries emerge. The semiconductor industry provides a prime example of this kind of trade. Brown and Linden (2009, pp. 87–91) wrote:

> Some chip companies with foreign design subsidiaries value the opportunity to design on a 24-h cycle because of the enormous pressure to reach the market ahead of, or no later than, competitors. One established US chip company adopted a rolling cycle between design centers in the United States, Europe, and India. More common is the bi-national arrangement used by a Silicon Valley start-up that had all of its design beyond the initial specification done by a China subsidiary established within months of the company's founding…. The Silicon Valley staff would review Beijing's work from the previous day, then spend up to three hours on the phone (starting around 5 pm California time) providing feedback and reviewing assignments for that day in Beijing. In a single-location firm this work-feedback cycle take two days instead of one.

Not only firms, but also consumers also prefer to consume services early taking the advantage of time zone differences. Ireland, pitching to host Europe's main international call centers, offers another example. Cairncross (1997, p. 219) emphasized the rise of the call-center service industry in Ireland, which is taking geographical advantage of being in between the U.S. and Europe.

To summarize above arguments: due to the communications revolution, time zone differences may become a primary driving force for service trade. Furthermore, these kinds of service trade invite new types of incentives for FDI. From home consumers'/firm's viewpoints, it is preferable that some subsidiaries locate at *distant* areas to serve the Home market. Although this point is at odds with the "proximity advantages" of FDI (e.g., Brainard 1997), it seems to be important to consider these new types of FDI incentives. Related to these phenomena, Marjit (2007) examined the role of international time zone differences in a vertically integrated Ricardian framework. It has been shown there that time zone differences emerge as an

[1]A substantial amount of empirical research has also emerged very recently revolving around the idea of time zones and trade. This further strengthens the underlying encouragement to write this paper. A representative sample of empirical papers consists of Anderson (2012), Christen (2012), Costinot et al. (2012), Dettmer (2011) etc.

independent driving force of international trade besides taste, technology, and resource endowment.[2]

What remains, however, unanswered is the relationship between firm productivity and FDI with time zone difference. Based on casual empiricism, we believe that time-saving technological improvement (e.g., utilization of communications networks such as the Internet) can trigger a series of events that leads to reallocations of industry structure via FDI. In the existing literature on FDI and firm heterogeneity, however, relatively few attempts have been made to address the role of time zone differences on FDI decisions.[3] This seems to suggest that the focus on "trade using different time zones" should be accompanied by a focus on firms' FDI decisions. Therefore, the main purpose of this study is to illustrate, with simple FDI model with heterogeneous firms, how a time-saving improvement in service trade using different time zones can have a huge impact on firms' FDI decisions.

For these purposes, based on Helpman et al. (2004), we propose a simple two-country model with heterogeneous firms that capture the role of FDI via utilizing time zone differences. Two countries (Home and Foreign) are assumed to be located in different time zones and there is no overlap in daily working hours. We further assume that both countries are small in nature. The key assumption of our model is that domestic service production requires two consecutive work days and that products are ready for sale after two workdays-domestic delivery bears significant costs in terms of delay.[4] In contrast to this, the utilization of communications networks allows production in a foreign country with nonoverlapping work hours, and service trade via networks enables a quick delivery and low shipping costs. In other words, imported services, whose production benefits from time zone differences, provide higher value than domestically produced service.

Based on the model outlined above, this study shows that productivity of the firms undertaking FDI is higher than the productivities of non-FDI firms. Although the results look quite similar with Helpman et al. (2004), the direction of service trade flow is totally different: foreign subsidiaries of high-productivity firms serve the Home market. In other words, in the sense of timeliness, building Foreign subsidiaries via FDI implies building subsidiaries closer to the Home market (see, Fig. 1). This result is in contradiction with the conventional wisdom that asserts why foreign subsidiaries of high productive domestic firms via FDI

[2]Jones et al. (2005) also emphasize the role of time zone differences as a determinant of efficient worldwide division of labor. Furthermore, fragmentation of production stages and of service provision has been studied within astatic trade-theoretic framework by Jones and Kierzkowski (1990), Grossman and Helpman (2005), Van Long et al. (2005), Do and Long (2008).

[3]In an important contribution, Helpman et al. (2004) show that the productivity of the firms undertaking FDI is higher than the productivity of the exporters. Following this, Mukherjee (2010) shows that the theoretical prediction of Helpman et al. (2004) may not hold. In addition, Helpman (2006) provided an excellent survey on the literature on FDI with heterogeneous firms.

[4]For related issues and modelling of such cost in time zones and trade context see Marjit (2007), Kikuchi et al. (2013), Mandal et al. (2014).

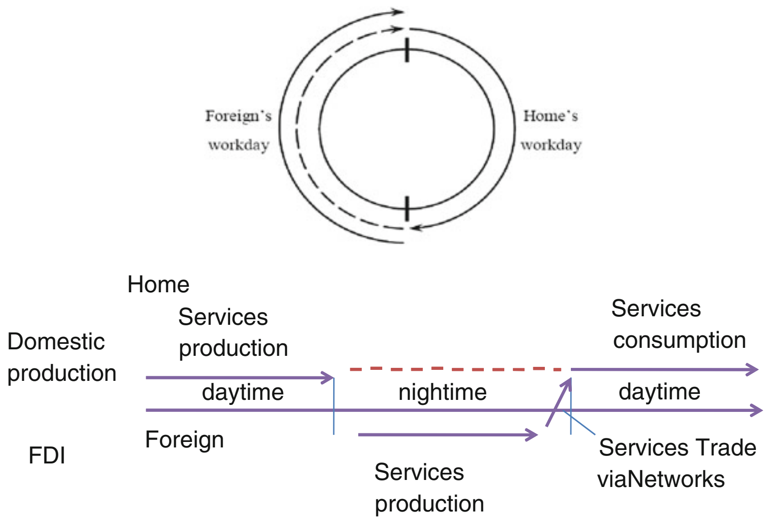


Fig. 1 Time zone exploitation and service delivery

serve the Foreign market. Whereas, in this paper, we primarily focus on how productivity of firms determines location of their production for serving their domestic market.

2 The Model and Basic Results

Suppose there are two countries, Home and Foreign, which are endowed with one factor of production (labor). They are located in different time zones and there is no overlap in daily working hours: when Home's daytime working hours end, Foreign's daytime working hours begin (Fig. 1).

There are two types of goods: a homogeneous good and a large variety of differentiated services. Only Home consumers demand the differentiated services, while both countries demand the homogeneous good.

The preference of the representative Home consumer is given by:

$$u = (1-\beta)\log z + \frac{\beta}{\alpha}\log\left(\int_{v}[x(v)]^{\alpha}\mathrm{d}v\right) \tag{1}$$

where z is the consumption of the homogeneous good, $x(v)$ is the consumption of variety v, $\alpha = \frac{(\varepsilon-1)}{\varepsilon}$, $\varepsilon > 1$ is the elasticity of substitution between varieties. Thus, we have the following demand function for x:

$$x(v) = A[p(v)]^{-\varepsilon} \tag{2}$$

$$A = \frac{\mathrm{BE}}{\int_0^n [p(v)]^{1-\varepsilon} \mathrm{d}v} \tag{3}$$

where E is the aggregate level of spending in Home, n is the measure of service varieties available in Home, and $p(v)$ is the consumer's price of variety v.

The homogeneous good is produced with constant returns, using labor as input. Units are chosen in such a way that one unit of labor produces one unit of output. As usual, no transport costs exist for the homogeneous good, which serves to tie down the wage rate. Also assume that the parameters of the model are such that both countries produce the homogeneous good. Thus, wages (hereafter set to unity) across countries are identical and constant.

Now, let us turn to the differentiated services. To simplify the analysis, we assume that the difference in productivities of firms exists only for Home firms. To enter the industry, a firm bears the fixed costs of entry f_{E}, measured in labor units. An entrant then draws a labor-per-unit-output coefficient a from a distribution $G(a)$. Upon observing this draw, a firm may decide to exit and not to produce. If it chooses to produce domestically, however, it bears additional fixed overhead labor costs f_{D}. On the other hand, if it chooses to serve the domestic (Home) market via foreign direct investment (FDI), it bears additional fixed costs f_{I} (e.g., build up communications networks between two countries). We assume

$$f_{\mathrm{I}} > f_{\mathrm{D}}. \tag{4}$$

The key assumption is that domestic production requires two workdays and that service is ready for sale after two workdays—the delivery of domestic product or service involves significant costs in terms of delay. In contrast to this, utilization of communications networks allows part production in Foreign country with nonoverlapping work hours, and trade via networks/Internet enables quick delivery. For these reasons, imported service products, whose production benefit from time zone differences provide higher value than domestically produced services.

In order to capture this point, we assume that shipment of products incurs the "iceberg" effect of delivery costs: to sell one unit of Foreign products in the Home market, $\tau(\tau > 1)$ units must be shipped. Thus, the price of the Foreign services becomes τ times higher than its original price. One can interpret τ as a measure of the inverse of the "delivery timeliness" of Foreign products in the *Home* market: a lower value of τ implies a quicker delivery.

As mentioned above, domestic production is ready for sale after two workdays, whereas imported services whose production benefits from time zone differences are available sooner (see Fig. 1). To parameterize the timing of delivery, we treat the utilization of communications networks (i.e., technological improvement) as a reduction in the delivery time of imported products (i.e., a decrease in).[5] Let us denote the Foreign services' delivery timeliness before technological change as τ_1 and that after change as $\tau_2 \cdot \tau$ represents the cost of communication which is required when TZ difference is exploited. Then the following condition holds[6]

$$\tau_1 > \tau_2 \geq 1 \tag{5}$$

Note that this effect comes not from lower production costs in Foreign, but from faster delivery. In other words, in a sense of timeliness, building Foreign subsidiaries via FDI implies building subsidiaries closer to the Home market (see, Fig. 1).

As noted above, preferences (1) generate a demand function $Ap^{-\varepsilon}$ for every brand of the service products, where the demand level A is exogenous from the point of view of the individual supplier. In this case, the brand of a monopolistic producer with labor coefficient a offers the price $p = \frac{\delta a}{\alpha}$ where $\delta > 1$ represents the loss of valuation from consumer's perspective due to untimely delivery and $1/\alpha$ represents the markup factor. So, essentially producers get a price equal to $\frac{p}{\delta} < p \cdot \delta$ gradually falls if consumers get the product early. This is a natural preference behavior of the consumers. As a result, the effective consumer price is $\frac{\delta a}{\alpha}$ for domestically produced services, and is $\frac{\tau_i a}{\alpha}$ for imported services.

Operating profits from domestic production for a firm with a labor output coefficient a is

$$\pi_D = a^{1-\varepsilon} B - f_{\mathrm{D}} \tag{6}$$

$$B = \frac{(1-\alpha)A}{\alpha^{1-\varepsilon}} \tag{7}$$

On the other hand, the operating profit from FDI (serving Home market via communication network) is

$$\pi_{Ii} = (\tau_i a)^{1-\varepsilon} B - f_I, \quad i = 1, 2 \tag{8}$$

These profit functions are depicted in Fig. 2. In this figure, $a^{1-\varepsilon}$ is represented on the horizontal axis. Since $\varepsilon > 1$, this variable increases monotonically with labor productivity $1/a$, and can be used as a productivity index. Two profit functions are increasing linear functions of this index. More productive firms are therefore more profitable in all these two activities.

The least productive firms expect negative operating profits and therefore exit the industry. This happens to all firms with productivity levels below $(a_{\mathrm{D}})^{1-\varepsilon}$. The slope

[5] τ depends only on ICT. ICT revolution reduces τ and opens up the possibility of utilizing TZ difference.

[6] In one extreme $\tau_2 = 1$ indicating zero communication cost.

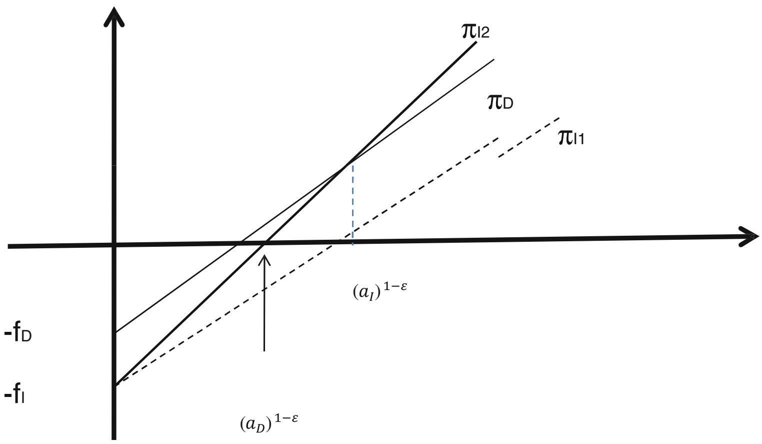


Fig. 2 Time zone exploitation and building foreign subsidiaries

of π_1 equals $(\tau_i)^{1-\varepsilon}B$, $i = 1, 2$, which depends on the technological condition of communication network (see (5)). When $\tau_1 > \delta > 1$, FDI is always unprofitable (a dotted line). If $\tau_1 > \delta > \tau_2 > 1$, firms with productivity above $(a_I)^{1-\varepsilon}$ gain more from FDI.[7] For this reason, given that $1 < \tau_2 < \delta$, firms with productivity levels between $(a_D)^{1-\varepsilon}$ and $(a_I)^{1-\varepsilon}$ choose domestic production while those with higher levels build subsidiaries in Foreign and produce Foreign services. In other words, via time-saving technological improvement, firms with higher productivity begin to build Foreign subsidiaries.

The (fixed) costs of building Foreign subsidiaries can be offset by a lower delivery (time) costs of services denoted by δ.

Proposition 1 *Given that* $1 < \tau_2 < \delta < \tau_1$, *firms with higher productivity choose to FDI and provide "Foreign" services for Home market.*

It is evident from the figure that the cutoff coefficients are determined by

$$(\delta a_D)^{1-\varepsilon}B = f_D \tag{9}$$

$$\left[\tau_2^{1-\varepsilon} - \delta^{1-\varepsilon}\right](a_I)^{1-\varepsilon}B = f_I - f_D \tag{10}$$

Free entry ensures equality between the expected operating profits of a potential entrant and the entry costs f_E. This condition can be expressed as

$$\left(\left[(\tau_2)^{1-\varepsilon} - \delta^{1-\varepsilon}\right]V(a_I) + V(a_D)\right)B - [G(a_I)(f_I - f_D) + G(a_D f_D)] = f_E \tag{11}$$

[7] τ_1 and τ_2 are the cost of communication in pre and post technological revolution phase, respectively with the condition that $\tau_2 < \tau_1$.

$$V(a) = \int_0^a y^{1-\varepsilon} \mathrm{d}G(y) \tag{12}$$

Equations (9)–(11) provide implicit solutions for the cutoff coefficients $a_\mathrm{D}, a_\mathrm{I}$, and the demand level B.

Combining (9) and (10), the following must hold

$$\frac{a_\mathrm{D}}{a_\mathrm{I}} = \tau_2 \left(\frac{f_\mathrm{I} - f_\mathrm{D}}{f_\mathrm{D}} \right) \left(\frac{\delta^{1-\varepsilon}}{\tau_2^{1-\varepsilon} - \delta^{1-\varepsilon}} \right)^{1/(\varepsilon-1)} \tag{13}$$

From (13), we can obtain the ratio of domestic production relative to FDI sales:

$$\frac{S_\mathrm{D}}{S_\mathrm{I}} = \frac{\int_{a_\mathrm{I}}^{a_\mathrm{D}} (\delta a)^{1-\varepsilon} B}{\int_0^{a_\mathrm{I}} (\tau_2 a)^{1-\varepsilon}} = \left(\frac{\delta}{\tau_2} \right)^{1-\varepsilon} \left[\frac{V(a_\mathrm{D})}{V(a_\mathrm{I})} - 1 \right] \tag{14}$$

In order to explore the effects of productivity dispersion on the ratio $\frac{S_D}{S_I}$, we parameterize $V(a)$ by parameterizing the distribution $G(a)$. For expositional purposes, let us use a Pareto distribution with the shape parameter k.[8] Then, we can obtain

$$V(a) = \int_0^a y^{1-\varepsilon} \mathrm{d}G(y) = c a^{k-(\varepsilon-1)} \tag{15}$$

where c is constant and it is assumed that $k > (\varepsilon - 1)$. Plugging back in (14), we can obtain

$$\frac{S_\mathrm{D}}{S_\mathrm{I}} = \left(\frac{\delta}{\tau_2} \right)^{1-\varepsilon} \left[\left\{ \left(\frac{f_\mathrm{I} - f_\mathrm{D}}{f_\mathrm{D}} \right) \left(\frac{\delta^{1-\varepsilon}}{\tau_2^{1-\varepsilon} - \delta^{1-\varepsilon}} \right) \right\}^{\frac{k-(\varepsilon-1)}{\varepsilon-1}} -1 \right] \tag{16}$$

It is then straight forward to see that the ratio of domestic production to FDI is decreasing in delivery timeliness of imported services as δ is assumed to be a constant because people's valuation for waiting time does not change very quickly. It is also decreasing in productivity dispersion, as parametrized by lower k.

Proposition 2 *A decrease in one country's delivery costs for imported services decreases the relative sales of domestic production. Also, an increase in productivity dispersion decreases the relative share of domestic production.*

[8]See, for example, Helpman et al. (2003, 2004).

Let us suppose that Home is a developed country, while Foreign is a developing country. Our result suggests that a time-saving technological change improvement in the developed country, which then requires more services provided with the benefit of time zone differences, triggers high-productivity firms' FDI toward the developing country. Jones and Marjit (2001) argue that, in a world in which the costs of service links are falling drastically, fragmentation of production process offers new opportunities to developing countries. The present result in FDI with high-productivity firms provides some theoretical grounds for such a development process.

3 Conclusions

In line with Helpman et al. (2004), here we have developed a two-country model to check if productivity determines the trade pattern in presence of nonoverlapping time zones between trading countries. It has been shown here that, even in absence of any wage differential, a fall in communication cost itself may trigger trade in finished and unfinished services. The pattern of trade we describe is different from what is explained in the existing literature. In the same line, we also explain why high productive firms opt for FDI in trading countries and import back the finished product. This argument does not naturally go with the traditional arguments for FDI and trade pattern where foreign subsidiaries serve the foreign market. Therefore, introduction of nonoverlapping time zones with low communication cost adds an interesting dimension to the FDI and trade pattern literature. Our results would be strengthened further if one introduces wage differential as predicted in Jones and Marjit (2001).

References

Anderson E (2012) Longitude matters II: time differences, communication and trade", Unpublished manuscript, School of International development, University of East Anglia

Brainard SL (1997) An empirical assessment of the proximity-concentration trade-off between multinational sales and trade. Am Econ Rev 87:520–544

Brown C, Linden G (2009) Chips and change: how crisis reshapes the semiconductor industry. MIT Press, Cambridge, MA

Cairncross F (1997) The death of distance. Harvard Business School Press, Boston, MA

Christen E (2012) Time zones matter: The impact of distance and time zones on services trade, Working Papers 2012–14, University of Innsbruck

Costinot A, Vogel J, Wang S (2012) An elementary theory of global supply chains. CESifo working paper series 3402

Dettmer B (2011) International service transactions: is time a trade barrier in a connected world? Jena Economic Research Papers 2011–003

Do V, Van Long N (2008) International outsourcing under monopolistic competition: winners and losers. In: Marjit S, Yu E (eds) Contemporary and emerging issues in trade theory and policy. Emerald Group, Bingly, UK, pp 345–366

Grossman GM, Helpman E (2005) Outsourcing in a global economy. Rev Econ Stud 72:135–159
Helpman E (2006) Trade, FDI, and the Organization of Firms. J Econ Lit 44:589–630
Helpman E, Melitz MJ, Yeaple SR (2003) Export Versus FDI. Working Paper No. 9439, National Bureau of Economic Research (Cambridge, MA)
Helpman E, Melitz MJ, Yeaple SR (2004) Export versus FDI with heterogeneous firms. Am Econ Rev 94:300–316
Jones R, Kierzkowski H (1990) The role of services in productions and international trade: a theoretical framework. In: Jones R, Kruger A (eds) The political economy of international trade. Blackwell, Oxford, pp 31–48
Jones R, Kierzkowski H, Lurong H (2005) What does evidence tell us about fragmentation and outsourcing? Int Rev Econ Finan 14:305–316
Jones R, Marjit S (2001) The role of international fragmentation in the development process. Am Econ Rev 91:363–366
Kikuchi T, Marjit S, Mandal B (2013) Trade with time zone differences: factor market implications. Rev Dev Econ 17(4):699–711
Van Long N, Riezman R, Soubeyran A (2005) Fragmentation and services. North Am J Econ Fin 16:137–152
Mandal B, Marjit S, Nakanishi N (2014) Time Zones, Factor Prices and Inflow of Educational Capital: Changing Sectoral Composition, mimeo, University at Albany-SUNY
Marjit S (2007) Trade theory and the role of time zones. Int Rev Econ Fin 16:153–160
Mukherjee A (2010) A note on firm-productivity and foreign direct investment. Econ Bull 39:2107–2111

MNEs and Export Spillovers: A Firm-Level Analysis of Indian Manufacturing Industries

Maitri Ghosh

Abstract FDI inflows, and hence MNE participation, in India during post-reforms have led to higher exports across sectors. In this context foreign firms might have an important role to play particularly in terms of export spillovers. This paper, in specific, estimates direct export spillovers from foreign firms to domestic firms across industries in Indian manufacturing during post-reforms. Heterogeneity of firms is controlled for in this study both in terms of labour productivity and specific costs. Dynamic panel data estimation results do not lend support to the evidence of export spillovers from foreign to domestic firms. It is evident in only a few industries. Rather factors like imported technology, both in embodied and disembodied form, marketing and distribution networks, credit availability and firm heterogeneity explain export intensity of domestic firms in Indian manufacturing during post-reforms.

Keywords MNEs · Export spillovers · Dynamic panel data estimation

1 Introduction

This paper empirically investigates the role of Multinational Enterprises (MNEs) in generating export spillovers on domestic firms across Indian manufacturing industries during post-reforms. MNEs, the major channel through which FDI flows into the host economies, often use the host country as an export platform and access

The author is thankful to K.L. Krishna, Sudip Chaudhuri, Sarmila Bannerjee, Saikat Sinha Roy, Ajitava Raychaudhuri and an anonymous referee for their comments on an earlier draft of the paper. The author is grateful to the Centre for Advanced Studies, Department of Economics, Jadavpur University, Kolkata, for allowing to use Prowess database used in this study. However, the author is solely responsible for errors, if any.

M. Ghosh (✉)
Bethune College, 181 Bidhan Sarani, Kolkata 700006, India
e-mail: maitri_ghosh@yahoo.com

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_3

foreign markets with ease. Hence, MNE operations which induce economic spillovers are likely to be export-enhancing for the host economy (Enderwick 2005). Such spillovers from the MNEs can take the form of export spillovers and/or technology spillovers which lead to productivity growth in the host economy (Caves 1972; Globerman 1979; Blomstorm and Perssion 1983; Haddad and Harrison 1993). With wide ranging reforms and growing FDI inflows as a consequence, it is critical to understand the role of MNEs. In particular, it is worthy to understand whether exporting activities of foreign firms have any significant impact on the export performance of domestic firms in Indian manufacturing, a phenomenon termed as direct export spillovers.

The literature explores horizontal and vertical spillovers from foreign to domestic firms. Horizontal spillovers mostly involve sector specific technical knowledge, while vertical spillovers are more general than sector-specific. The spillover benefits to the domestic firms arise through various channels of transmission[1] like imitation, skill acquisition, competition and exports. While the theoretical literature suggests imitation as the major spillover channel, Haacker (1999) and Fosfuri et al. (2002) argue that the knowledge that workers bring with them is the most effective spillover channel. Wang and Blomstorm (1992), Glass and Saggi (2002) emphasize on the role of competition for spillovers to be effective. The literature has also shown that FDI, new technology generation and technology transfer has often determined economic growth (Saggi 2002). MNEs form one of the major channels of technology spillovers to host countries, import of foreign technical expertise and their technology development (Findlay 1978; Das 1987; Wang and Blomstrom 1992; Perez 1997). Such spillovers might occur through the potential channels of demonstration effects, labour turnover or vertical linkages.

Again, export as a major spillover channel is also evident in the literature, with evidence on direct and indirect channels of spillovers. The direct approach relates to spillovers that are directly linked to foreign presence, while the indirect approach to the theory of spillovers explains some important channels of spillovers, namely competition, demonstration and imitation and information externalities. MNEs are believed to possess strong distribution networks, good infrastructure and linkages, knowledge of the consumers' tastes and preferences and regulatory arrangements in overseas markets. Domestic firms learn to export from MNEs through collaboration or imitation. Export activities of MNEs often produce externalities that enhance the export prospects of domestic firms (Rhee and Belot 1990). Aitken et al. (1997) show spillovers from information externalities with MNEs, but not from general export activity. Sousa et al. (2000) show positive spillovers from MNEs to domestic firms in the host economy through export information externalities, increased competition in the domestic market and demonstration effect. Kneller and Pisu (2007) also find significant export spillovers from the operations of foreign affiliates in the UK. Swenson and Chen (2014) show that own-industry multinational firm contact associates with more frequent, higher valued and long-lasting exports which

[1]See, for instance, Bergman (2006).

arise from beneficial spillovers. Ruane and Sutherland (2004) also find positive effects from the MNEs on both decisions to export and export share. Buck et al. (2007) find evidence of export spillovers across industries. Buckley et al. (2002) also show spillovers to Chinese local firms through transfer of marketing skills from MNEs. On the contrary, Barrios et al. (2003) find no evidence of MNE effect on the export share in Spain. Konings (2001) also suggest no positive spillover from foreign to domestic firms in Bulgaria, Romania and Poland. However, when absorptive capacity is considered, positive spillovers are found in R&D intensive firms in Bulgaria and Poland.

The studies reviewed above bring forth the evidence on MNE presence and export spillovers are far from conclusive. Further, with the exception of China, there are not many studies in particular for the emerging market economies including India. Further, spillover is mostly studied at the industry level and the results arrived at the industry level might not hold at the firm level, given heterogeneity of firms across industries *a la* Melitz (2003). In the Indian context however, only Chiara and Sasidharan (2010) carry out a firm-level analysis for the period 1994–1996 to explore the spillover channels. There still remain reasons to claim that there is little evidence of export spillovers from export orientation of foreign firms in India. This paper fills in the gap of identifying direct export spillovers at the firm-level after having controlled for firm heterogeneity across manufacturing sectors in India.

The paper is organized as follows. Section 2 puts forth some stylized facts on the FDI-export nexus in India. Section 3 discusses the analytical framework, the empirical model and method, and the database. Section 4 presents the empirical results. Section 5 summarizes the major findings of the paper.

2 FDI and Exports in India: Some Stylized Facts

The existing literature suggests a theoretical possibility and an empirical connect between MNE presence and export performance. In India, along with increasing FDI inflows, firm-level average export intensity[2] across manufacturing sectors increased during post-reforms, especially after the year 2000. The average export intensity for all the sectors, as is evident in Table 1, increased from 0.10 in 1990s to 0.15 in 2000s, the corresponding ratio for chemicals, transport equipment, machinery, food and beverage, textile and basic metal industries also increased after 2000. Here it is important to understand whether export performance significantly varies across ownership pattern,[3] given the perception that foreign firms perform better than their domestic counterparts especially during post-reforms.

[2]Export intensity at the firm level is calculated as the ratio of export of goods to sales expressed in percentage.

[3]Ownership in this analysis is not related to equity shares, as the data on equity for firms are not available for the entire time period under consideration.

Table 1 Firm-level average export intensity in India during post-reforms

	Food and beverage	Textiles	Chemicals	Basic metals	Machinery	Transport equipment	All industries
All sample firms							
1990s	0.24	0.22	0.09	0.08	0.04	0.10	0.10
2000s	0.28	0.29	0.18	0.15	0.10	0.11	0.15
Only domestic firms							
1990s	0.06	0.21	0.06	0.53	0.04	0.01	0.14
2000s	0.07	0.33	0.12	0.01	0.06	0.06	0.28
Only foreign firms							
1990s	0.04	0.02	0.04	neg	0.05	neg	0.11
2000s	0.03	0.03	0.16	neg	0.07	0.0005	0.28

Note 'neg' refers to negligible
Source Calculations based on CMIE, PROWESS database

Table 1 shows that average export intensity has increased for domestic firms and foreign firms alike post 2000. The increase in particular is large for chemical industry. For the domestic chemical firms the average export intensity doubled from 0.06 to 0.12 during post-2000, while for foreign chemical firms, average export intensity more than doubled from 0.04 to 0.16 during the same period. Similar trend is noticed in the other sectors including machinery, textile and transport equipment. The only exception to this pattern is in basic metals industries. Export intensity of foreign firms in food and beverages industry also shows a marginal decline after 2000. These findings of no better performance of foreign over domestic enterprises in India are in conformity with that of Kumar (1990), Pant (1993), Kumar and Siddharthan (1994) and Aggarwal (2002).

Prima facie, with improving export performance of domestic firms over the years along with that of foreign firms, the role of export spillovers from foreign to domestic firms becomes important. This paper investigates into the role of exporting activity of foreign firms in explaining the international competitiveness of domestic firms.

3 Analytical Framework

MNEs, as evident in the literature, not only bring in capital investment but also prove to be beneficial to the host country as they possess product/process technology, marketing and managerial skills, etc., and use the host country as export platform. MNEs can also potentially induce domestic firms to export (Kumar 1994). These perspectives on MNEs help develop the framework for firm-level export performance in an emerging market economy. The following model is essentially a supply side one based on Aitken et al. (1997). The profit function of a representative firm is as follows:

$$\Pi = P_d Q_d + P_f Q_f - C(Q_d + Q_f) - S \quad (1)$$

$$\text{s.t.} Qd, Qf > 0$$

The subscripts d and f are for the domestic and the foreign markets respectively. The cost in this framework is as follows:

$$C = C(h(Q_d + Q_f), j(Q_d + Q_f)) \quad (2)$$

The $h(.)$ function signifies the production costs. As production cost is independent of the market the commodity is sold, h is a function of both Q_d and Q_f. Again, apart from being productive, a representative firm bears certain costs to stay in the market. For instance, a firm incurs expenditure on R&D, advertising, creating export infrastructure, and developing market channels. This is represented by $j(.)$. Further, in order to export a firm has to bear certain costs which are sunk in nature and hence cannot be recouped. For example, these costs include learning about bureaucratic procedures and adapting products in the foreign market. Again, in each period to maintain presence in foreign market, minimum freight charges, insurance charges, costs of monitoring foreign customs procedure, etc., are to be borne by the firm (Das et al. 2007). Sunk cost is represented by S in the model. The cost function in Eq. (2) is specified as follows:

$$C(Q) = a/2(Q)2 + g[h(Q), j(Q)],$$

where a and g are scalar parameters.

Thus, rewriting the profit function we have:

$$\Pi = P_d Q_d + P_f Q_f - a/2(Q_d + Q_f)2 - g[h(Q_d + Q_f), j(Q_d + Q_f)] - S$$

Maximizing profit with respect to Q_f and setting $\delta\Pi/\delta Q_f = 0$, we get:

$$Q_f^* = [P_f - aQ_d - \delta g/\delta h[h'(.)] - \delta g/\delta j[j'(.)] \quad (3)$$

Similarly, setting $\delta\Pi/\delta Q_d = 0$, we get:

Hence,

$$Q_d^* = [P_d - aQ_f - \delta g/\delta h[h'(.)] - \delta g/\delta j[j'(.)] \quad (4)$$

Exports of the representative firm thus essentially depend on P_f, P_d, h and j. There can be other supply side factors like size and age of the firm, productivity, import of technology and firm ownership which are important factors in export determination. Now, with India being assumed to be a small country, P_f is can be said to be given across all firms in each sector. Again, P_d is also assumed to be same for all firms producing similar products. In the analysis that follows, P_f, P_d and hence their impact are not taken into consideration. In what follows is a detailing out of various other supply side factors including firm-specific characteristics, specific costs, technology and credit availability that are controlled for in this study.

In the empirical literature, as Aw et al. (2000) and Clerides et al. (1998) show that a few productive firms, which expect a profit stream sufficiently high to cover the sunk costs of entry into a foreign market, find it profitable to export. Following Melitz (2003), models explicitly postulate that firms are heterogeneous based on productivity and only productive firms self select into export markets. In order to incorporate heterogeneity of firms in our model, following the literature we use firm productivity. In this study, productivity is measured using labour productivity instead of total factor productivity, as is done in the 'heterogeneity' literature.[4]

Exploring a foreign market requires strong marketing and distribution networks. If a firm incurs expenditure on advertisement marketing and distribution, and creates service networks, it might attain cost competitiveness in exporting its product in a foreign market. These costs are sunk in nature and cannot be recovered (Baldwin 1999). In India expenditure on advertising, marketing and distribution however show a wide variance among Indian firms. Srinivasan and Archana (2011) is the only study accounting for such firm-specific costs while identifying factors determining exports.

In an increasingly knowledge-based world, technological capacity is seen as an important component of a country's international competitiveness and growth (Kumar and Aggarwal 2005). It is found in the literature that in-house research and development makes a firm cost competitive and thereby its export performance improves (Fargerberg 1988; Soete 1981). Veugelers and Cassiman (1999), Roper and Love (2002), Aggarwal (2001), Kumar and Siddharthan (1997), Patibandala (1995) provide evidence on R&D expenditure having significant positive impact on export intensity. Further, import of technology, embodied and disembodied is one of the major channels of technology acquisition by firms. Embodied technology is imported in the form of raw materials, intermediate goods and capital goods, while imported disembodied technology includes patented knowledge, technical know-how, drawings and designs, etc. It is believed that in-house R&D and technology imports makes a firm cost competitive and thereby induces exports to grow. In this analysis four variables, namely, in-house R&D expenditure, import of raw materials, import of capital goods, foreign technical know-how import are taken into account while controlling for technology acquisition by firms.

Large sized firms are expected to have greater resource base and better risk perception of the international market and hence aid exports. However, Bernard and Jensen (2004) perceive size to be a proxy for several effects including economies of

[4]Srinivasan and Archana (2011) have also used labour productivity in the estimation by arguing that firms can gain international competitiveness through lower cost arising out of higher labour productivity indicating products being produced by quality labour.

scale that determines the export attitude and performance of a firm (Kumar and Pradhan 2003). Age of a firm, in the literature, also shows the extent of a firm's learning experience leading to greater experimental and tacit knowledge (Bhaduri and Ray 2004). Older firms, with experience in exporting are able to bear the sunk costs of exporting and have better capability to export (Rasiah 2003; Iyer 2010).

There are empirical studies, which explain the impact of credit availability on firm's export performance.[5] Within the Melitz (2003) framework, Chaney (2005), Helpman et al. (2008) and Manova (2008) show that in addition to heterogeneity of firms in terms of productivity and capability to overcome sunk costs, credit constraints also affect exports of firms. In the Indian context, Kapoor et al. (2011) have established a causal link from availability of subsidized credit to small firms' credit constraints to real outcomes of exporting firms.

On the whole, the literature suggests that exporting activity of a firm can be explained by various firm-specific factors given heterogeneity of firms. This paper explores the determining factors explaining export intensity of domestic firms in Indian manufacturing across sectors during post-reforms. In doing so the role of export spillovers from foreign firms to domestic firms and firm heterogeneity explained in terms of labour productivity and sunk costs is understood. This is done controlling for factors like age, size, R&D intensity, import of foreign technology and credit availability of firms.

3.1 The Estimation Model

In the estimable form, export intensity of a domestic firm is found to depend on exporting activity of foreign firms and other firm-specific supply side factors including age, size, technology imports and credit availability and heterogeneity measured in terms of productivity. The estimation model is as follows:

$$\begin{aligned}\mathrm{Domx}_t =\ & \alpha_0 + \alpha_1(\mathrm{Domx}_{t-1}) + \alpha_2(\mathrm{age}_t) + \alpha_3(\mathrm{size}_t) + \alpha_4(\mathrm{rdi}_t) + \alpha_5(\mathrm{impr}_t) \\ & + \alpha_6(ki_t) + \alpha_7(\mathrm{fptr}_t) + \alpha_8(\mathrm{mkt\,cos}\,t_t) + \alpha_9(\mathrm{pdtivity}_t) + \alpha_{10}(\mathrm{crdt}_t) + \alpha_{11}(\mathrm{FOR}_t) \\ & + \mathrm{uit}\end{aligned} \quad (5)$$

where $\alpha_{i,\ i=1 \text{ to } 11} > 0$, and $\alpha_{11} > 0$ indicates positive export spillovers;

Domx	Export intensity of domestic firms
FOR	Average Export intensity of foreign firms
age	Absolute age of the firm in number of years
size	Ratio of firm sales to industry sales
rdi	Ratio of R&D expenditure to sales
impr	Ratio of expenditure on imports of raw materials to sales

[5]See, for instance, Mirabelle (2008) for Belgian firms, Greenaway et al. (2007) for UK firms and Paravisini et al. (2011) for Peruvian firms.

ki	Ratio of expenditure on imports of capital goods to sales
fptr	Ratio of technical fees and royalties paid abroad to sales
mktcost	Ratio of summed up advertising expenditure, marketing expenditure and distribution expenditure to sales
pdtivity	Ratio of value of output to salaries and wages
crdt	Ratio of total firm borrowing to value of output

Estimation of Eq. 5 requires detailing on the estimation method, correct measurement of these variables, and some details on the database.

3.2 The Method and Data

In our analysis we have used Dynamic Panel Data estimation technique. The Ordinary Fixed and Random Effect estimation methods are initially used to identify the control variables. To estimate time series and cross section data in a single equation framework, Panel data estimation technique is widely used in literature (see Baltagi 2001). It helps to simultaneously accommodate large volume of data set across time and distinguishes between time series movement and cross-sectional movement of the data. Dynamic effect can be examined in panel data analysis by introducing lagged dependent variables in the set of explanatory variables. The model, with one year lagged dependent variable, looks like:

$$Y_{it} = X_{it}'B_1 + Y_{it-1}B_2 + E_{it}$$

where $i = 1, 2, 3, \ldots., m$; $t = 1, 2, 3, \ldots, T$.

m = number of cross-sectional units; T = number of time period.

Here the lagged dependent variable, Y_{it-1} captures the entire historical impact of the explanatory variables. The problem however arises at the time of estimation. The Least Square Dummy Variable (LSDV) method and Feasible Generalised Least Square (FGLS) methods are inappropriate to estimate the model. Dynamic panel data estimation is usually carried out using ‘Generalised Methods of Moments’ (GMM). This is done by estimating the model in first difference to avoid the problem of endogeneity arising due to the presence of lagged endogenous variable in the set of explanatory variables. The GMM IV estimation of Arellano and Bond (1991) is applied to obtain unbiased consistent estimators. A 2-stage iteration method is used to get Arellano and Bond 2-step estimators. In order to obtain original Arellano and Bond estimates, no correction for the degree of freedom is carried out. In this type of estimation, Sargan test of overidentifying restriction is checked.

Despite limitations, firm-level data across sectors are obtained from Prowess Database published by the Centre for Monitoring Indian Economy (CMIE) for the period 1991–2010. This database is well suited for our analysis as PROWESS data is a panel of firms unlike the unit level ASI (Annual Survey of Industries) data where the identification of cross section is not revealed. PROWESS provides information from audited financial statements and thereby uses company balance sheets and income statements as sources of information. The database covers both listed and unlisted firms from a wide cross section of manufacturing, services, utilities and financial industries covering 60–70 % of organized sector in India, 75 % of corporate taxes and 95 % of excise duties collected by the Government of India (Goldberg et al. 2010).

In this study, information on only manufacturing firms is used. The sectors include chemicals, machinery, transport equipments and food and beverages industries. Textiles and metal and metal products are excluded as fewer foreign firms exist for these sectors in the sample dataset. The dataset includes a cross section 1238 domestically owned firms for chemicals, 707 firms for machinery, 315 firms for transport equipments and 55 firms for food and beverages industries. Firms are categorized according to ownership patterns, domestic and foreign enterprises. PROWESS provides data for foreign promoter's equity holdings. If for a company, equity holding of the foreign promoter exceeds 25 %, it is classified as a foreign owned firm or a "FDI firm". However, PROWESS reports data on foreign promoter's equity holdings only for post 2001 period. As this study covers a twenty year period (1991–2010), the information on equity holdings to identify company ownership could not be used. Further, numerous missing values of equity participation also do not auger well with the empirical analyses being carried out. However, the database provides separate information on the ownership group of firm in the sense of whether a firm is 'Private Indian', 'Private Foreign' or a 'State-run' enterprise, etc. This information is used in the study to identify domestic and foreign ownership of firms. While estimating export spillovers it is postulated that the positive impact of foreign firms on the exporting behaviour of the domestic firms is indicative of export spillovers from foreign to domestic enterprises. As specified in Eq. (5), average export intensity of the foreign firms is one of the variables determining the domestic firm's export intensity. Dynamic panel data estimation is carried out. However, for four separate industries, which by nature are widely different, the model specification has been changed depending on the industry characteristics. Sargan test and Arrelano–Bond (2) tests show that the instrumental variables used in dynamic panel data estimation are valid and are not correlated with the error terms in both the specifications. The model estimated is thus identified. The export spillover results are presented in Table 2.

4 Empirical Results

The estimation results presented in Table 2 provides evidence on the significant positive effect of exports of foreign firms, FOR, on domestic firms' exports in India's transport equipment industries. This suggests presence of export spillovers from foreign to domestic firms in this industry. For most other industries, FOR

Table 2 Firm-level export spillovers: a dynamic panel data analysis

	Chemical	Machinery	Transport equipment	Food and beverages
$Domx_{t-1}$	**0.472*** (10.85)	**0.715*** (62.02)	**0.067*** (6.50)	−0.05 (−1.52)
FOR_t	0.16 (0.8)	0.04 (0.73)	**1.15*** (11.52)	**−0.627**** (−2.29)
Age_t	**0.423*** (3.77)	**−0.09***** (−1.73)	**0.80*** (37.48)	**−2.31*** (−6.43)
$Size_t$	40.76 (0.54)	**−0.001*** (−2.71)	22.5 (1.12)	**6.94*** (2.78)
Rdi_t	**−75.3***** (−1.79)	**98.41**** (2.15)	**224.56***** (1.72)	**1330.83*** (2.99)
Rdi_t^2	**23.9***** (1.79)	**−654.48**** (−2.35)	**–45477.33*** (−3.81)	
$Impr_t$	**16.3*** (5.89)	**−0.01*** (−2.91)	**−15.7*** (−10.00)	**−170.34**** (−1.87)
$Impr_t^2$	**−1.97*** (−5.90)	**0.0003*** (2.89)	**19.10*** (8.58)	**758.56**** (1.69)
Ki_t	−10.4 (−0.77)	**2.21**** (1.81)	**10.39*** (4.20)	**138.03*** (2.82)
Ki_t^2	14.7 (0.82)		**−6.29*** (−3.27)	
$Fptr_t$	−45.73 (−1.63)	−119.7 (−1.47)	**266.57*** (8.48)	**−7202.4**** (−2.20)
$Fptr_t^2$	1.12 (1.64)	**1717.9***** (1.86)	**−416.93*** (−8.42)	**136846.1**** (1.96)
$Mktcost_t$	**116.6*** (3.96)	**9.57*** (1.48)	**41.02*** (5.16)	**181.12*** (4.33)
$Mktcost_t^2$	**−374.4*** (−3.29)	**−34.75* (−1.11)**		
$Pdtivity_t$	**0.036**** (2.35)	**−0.039*** (−5.79)	**0.090*** (5.01)	**0.025*** (3.81)
$Pdtivity_t^2$	**−0.00001**** (−2.24)	**0.00002*** (5.28)		
$Crdt_t$	**1.32*** (3.44)	**−0.030*** (−3.52)	**0.018**** (1.91)	**0.044*** (3.72)
$Crdt_t^2$				**−0.0001*** (−3.93)
Sargan test Chi Square	28.46	33.41	38.35	15.88
AR(2) *Z* value	−0.10	−0.87	−1.58	0.44
Number of observations	1238	707	315	55

Note *z* values are provided in parentheses
*1 % level of significance, **5 % level of significance, ***10 % level of significance

remains positive though insignificant. This implies that presence of foreign firms do not necessarily augment exports of domestic firms in Indian manufacturing. Interestingly, for food and beverages, a significant negative spillover coefficient is observed. More value-added products like processed and packaged food, marine products, etc., constitute the export basket of food items post-reforms. With regard to export of these items, foreign firms are able to meet international quality standards while domestic firms are often not in a position to meet the international quality standards for food items required for exports.[6] Hence, foreign spillover effects do not play an important role in this sector as well. Almost similar results are

[6]Jongwanich (2009) show that stringent SPS standards prevalent in the processed food market globally often deter such exports from developing countries those are unable to meet the prevalent standards.

found by Aitken and Harrison (1999) suggesting that foreign ownership has significant negative productivity spillovers on domestically owned firms. However, estimation results suggest a significant path dependence across most sectors as the lagged dependent variable turns out to be positive and significant for all sectors barring food and beverages. This is an expected result as the exporting behaviour of the previous period is likely to affect the exporting behaviour of firms in the next period. Large size of firms, on the other hand is crucial for exporting of the domestic firms in the food and beverage sector. This is however not true for the other sectors. In particular, smaller firms are likely to export for machineries.

Factors like in-house R&D, marketing and distribution costs and import of raw materials explain the export performance of the domestic firms. As MNEs are equipped with better technology, they impact on domestic firms to remain competitive. Thus competition from foreign affiliate can have positive spillover effect on the domestic firms. Hence, upgradation of existing technology becomes important for the domestic firms to face competition from MNEs. This may be one of the reasons why R&D significantly explains the export intensity of the domestic firms across sectors. However, there are other factors as well that determine exports of domestic firms. Exporting requires very strong distribution networks, good infrastructure and linkages, knowledge of the consumers' tastes and preferences, regulatory arrangements in overseas markets, etc. Investment in advertising, marketing and distribution gives Indian domestic firms in the food and beverage, chemicals, machinery and transport equipment industries edge in exporting. In this study advertisement, marketing and distribution cost explain the sunk cost incurred to penetrate in the foreign market. The results conform to the theoretical conjecture that firms are heterogeneous in terms of sunk costs and the capability of overcoming this sunk cost of entering a foreign market is quite an important factor to explain export intensity. Srinivasan and Archana (2011) as well arrived at similar results.

Again, older domestic firms learn to export from the MNEs. Older firms with high productivity are better performers. The estimation results show that age of the firm, measured in terms of number of years in operation since inception, plays a significant role in determining firm-level export performance of the high technology industries like chemicals and transport equipments. This suggests that older firms have acquired the capability to penetrate the world market particularly for the high technology sectors. However, for firms in industries like machinery and low technology sector like food and beverages, the relationship between age and export intensity significantly negative. The older firms in the machinery sector which started operations during import substitution largely cater to the domestic market. This pattern is despite wide ranging reforms. The significantly negative relationship between age and export intensity in the food and beverage industry is in conformity to the findings of Kumar and Pradhan (2003) suggesting that older firms in low technology industries concentrate more on the domestic market during post reform.

For all major industries, heavy dependence on the imported raw materials is found to be a major driving force. Import of raw materials, capital goods and foreign technical knowhow by firms is one of the major sources of acquiring knowledge from rest of the world and in achieving cost competitiveness by using frontier technology and cheaper inputs. This impacts export intensity of domestic firms positively. Disembodied foreign technology aids the process. For chemicals, machinery, transport equipment and textiles, import of raw materials has significant positive impact on firm-level export intensity. This is as per expectations as most industries are knowledge-based and they crucially depend on imported raw materials to be globally competitive. A significant non-linear relationship exists between import of raw materials and export intensity across sectors. Import of capital goods is another important way to bring in global frontier knowledge in embodied form, while import of foreign technology brings in foreign design, technological expertise and knowledge in disembodied form. It is evident from the above results that import of capital good has a positive impact on export performance of all sectors except chemicals. Further, barring chemicals import of technology in disembodied form also plays a significant role in explaining export performance across sectors.

The existing literature shows that firm productivity is one of the major ways to explain firm heterogeneity. Not only labour productivity across sectors improved during post-reforms in 1991, Parameswaran (2014) show significant firm-level total factor productivity growth across most industries since the early 1990s, with a significant step up in productivity growth in food products, textiles, chemicals, basic metals and metal products, machinery—non-electrical, electrical and electronic—and transport equipment industries since 2000–01.[7] Estimation results as shown in Table 2 reflect that productivity of firms is significant in explaining the variations in firm-level export intensity in all major industries. Such relationship conforms to the pattern as shown in the theoretical conjectures by Melitz (2003). This is an expected result given the fact that there has been a significant improvement in firm-level productivity growth across industries since the early 1990s. Credit availability significantly explains the exporting behaviour of most Indian manufacturing. The trade-finance linkage empirically suits well for Indian manufacturing. We also find the presence of non-linearity in the relationship between credit availability and export performance in food and beverages.

4.1 Robustness Check for the Export Spillover Model

A caveat in the model as specified in (5) is that while the export intensity of the domestic firms varies across cross section and over time, the foreign export intensity being an average varies over time but remains constant across cross-section. Hence we check the robustness of the model by testing the model into two steps as follows:

[7]Similar results on firm-level productivity growth are arrived by Krishna and Mitra (1998), Balakrishnan et al. (2000), Balakrishnan et al. (2006), Topalova and Khandelwal (2011). Ghose and Roy Biswas (2014) shows that improvements in manufacturing productivity, though of a different period, are largely on account of technical change.

Table 3 Firm-level export spillovers: residual analysis

	Chemical	Machinery	Food and beverage	Transport equipment
FOR	0.05 (0.54)	−0.07 (−0.39)	**−0.54*** (−45.19)	**1.41*** (281.49)
Sargan test Chi Square	19.02	59.19	6.64	63.85
AR (2) *z* value	−0.08	−1.30	1.06	−1.22

Note z values are provided in parentheses
*1 % level of significance

$$\text{Dexpi}_t = \mu_0 + \mu_1(\text{Age}_t) + \mu_2(\text{size}_t) + \mu_3(\text{rdi}_t) + \mu_4(\text{impr}_t) + \mu_5(\text{ki}_t) + \mu_6(\text{fptr}_t) + \mu_7(\text{mkt cost}_t) + \mu_8(\text{Pdtivity}_t) + \mu_9(\text{Crdt}_t) + + \theta it \quad (6)$$

where $\mu_{i,\ i=1\text{ to }9} > 0$, and

$$\text{Dexpi}_t - \widehat{\text{Dexpi}}_t = \alpha_0 + \alpha_1(\text{FOR}_t) \quad (7)$$

After controlling for all the explanatory variables in (6), estimation results of (7) as revealed in Table 3 show no difference. In sum, foreign export spillovers continue to remain positive and significant only for the transport equipment industry. Significant negative spillover is observed for food and beverages. This is suggestive of robustness of the export spillover results.

5 Conclusions

Even though there can be several possible spillover channels, we investigated into direct export spillovers from MNEs to domestic firms in India. Dynamic panel data estimation results show significant export spillovers only in the transport equipment industry. Except in case of transport equipment industry, the channels of transmission from MNEs do not seem to be effective. Importantly enough, there is evidence of negative spillovers from foreign firms in case food and beverage exports. Rather, the capability to cover sunk costs by domestic firms, productivity, credit availability, local R&D and imported technology both in embodied and disembodied form turn out to be very significant in explaining exports of domestic firms. Direct export spillovers from foreign enterprises do not have that important a role to play in explaining the export performance of the domestic firms across most sectors.

References

Aggarwal A (2002) Liberalisation, multinational enterprises and export performance: evidence from Indian manufacturing. J Dev Stud 38(3):119–137

Aggarwal A (2001) Technology policies and acquisition of technological capabilities in industrial sector: a comparative analysis of the Indian and Korean experience. Sci Technol Soc 6(2):255–304

Aitken B, Harrison A (1999) Do domestic firms benefit from direct foreign investment? Evidence from venezuela. Am Econ Rev 89(3):605–618

Aitken B, Hanson GH, Harrison AE (1997) Spillovers, foreign investment and export behavior. J Int Econ 43(1–2):103–132

Arellano M, Bond S (1991) Some tests of specification for panel data: monte carlo evidence and an application to employment equations. Rev Econ Stud 58(2):277–297

Aw BY, Chung S, Roberts MJ (2000) Productivity and turnover in export market: micro-level evidence from the republic of Korea and Taiwan (China). World Bank Econ Rev 14(1):65–90

Balakrishnan P, Pushpangadan K, Suresh Babu M (2000) Trade liberalisation and productivity growth in manufacturing: evidence from firm-level panel data. Econ Polit Wkly 35(41):3679–3682

Balakrishnan P, Pushpangadan K, Parameswaran M, Suresh Babu M (2006) Liberalisation, market power and productivity growth in Indian industry. J Policy Reform 9(1):55–73

Baldwin R (1999) Agglomeration and endogenous capital. Eur Econ Rev 43(2):253–280

Baltagi BH (2001) Econometric analysis of panel data, 2nd edn. John Wiley and Sons, New York

Barrios S, Gorg H, Strobl E (2003) Explaining firms' export behaviour: R&D, spillovers and the destination market. Oxford Bull Econ Stat 65(4):475–496

Bergman A (2006) FDI and spillover effects in the indian pharmaceutical industry. RIS Discussion Paper 113. Research and Information System for Developing Countries, New Delhi

Bernard AB, Jensen JB (2004) Why some firms export? Rev Econ Stat 86(2):561–569

Bhaduri S, Ray AS (2004) Exporting through technological capability: econometric evidence from India's pharmaceutical and electrical/electronic firms. Oxf Dev Stud 32(1):87–100

Blomström M, Perssion H (1983) Foreign investment and spillover efficiency in an underdeveloped economy: evidence from the Mexican manufacturing industry. World Dev 11(6):493–501

Buck T, Filatotchev I, Wright M, Norlan P (2007) Different paths to economic reform in Russia and China: causes and consequences. J World Bus 35(4):379–400

Buckley JP, Jeremy C, Wang C (2002) The impact of inward FDI on the performance of China's manufacturing firms. J Int Bus Stud 33(4), 637–400

Caves R (1972) American industry: structure, conduct. Performance, New Jersey Prentice Hall

Chaney T (2005) Liquidity constrained exporters. National Bureau of Economic Research, Cambridge

Chiara F, Sasidharan S (2010) MNEs, technological efforts and channels of export spillover: an analysis of Indian manufacturing industries. Econ Syst 34(3):270–288

Clerides S, Lach S, Tybout J (1998) Is learning by exporting important? Micro-dynamic evidence from Colombia, Mexico and Morocco. Q J Econ 113(3):903–948

Das M, Roberts M, Tybout J (2007) Market entry costs, producer heterogeneity and export dynamics. Econometrica 75(3):837–874

Das S (1987) Externalities and technology transfer through multinational corporations: a theoretical analysis. J Int Econ 22(1):171–182

Enderwick P (2005) Attracting 'Desirable' FDI: theory and evidence. Transnational Corporations 14(2):93–119

Fagerberg J (1988) International competitiveness. Econ J 98(391):355–374

Findlay R (1978) Some aspects of technology transfer and direct foreign investment. Am Econ Rev 68(2):275–279

Fosfuri A, Motta M, Ronde T (2002) Foreign direct investment and spillovers through workers' mobility. J Int Econ 53(1):205–222

Ghose A, Roy Biswas P (2014) 'Impact of trade policy on manufacturing productivity in India: evidence using a non-parametric approach'. In Kathuria V, Rajesh Raj SN and Sen K (eds) Productivity in Indian manufacturing-measurements, methods and analysis. Routledge, New Delhi

Glass AJ, Saggi K (2002) Multinational firms and technology transfer. Scand J Econ 104(4):495–513

Globerman S (1979) Foreign direct investment and 'spillover' efficiency benefits in Canadian manufacturing industries. Canadian J Econ 12(1):42–56

Goldberg P, Khandelwal A, Pavcnik N, Topalova P (2010) Multiproduct firms and product turnover in the developing world: evidence from India. Rev Econ Stat 92(4):1042–1049

Greenaway D, Guariglia A, Kneller R (2007) Financial factors and exporting decisions. J Int Econ 73(2):377–395

Haacker M (1999) Spillover from foreign direct investment through labour turnover: the supply of management skills. CEP Discussion Paper. London School of Economics, London

Haddad M, Harrison A (1993) are there positive spillovers from direct foreign investment? Evidence from panel data for morocco. J Dev Econ 42(1):51–74

Helpman E, Melitz MJ, Rubinstein Y (2008) Estimating trade flows: Trading partners and trading volumes. Q J Econ 123(2):441–487

Iyer K (2010) The determinants of firm-level export intensity in New Zealand agriculture and forestry. Econ Anal Policy 40(1):75–86

Jongwanich J (2009) Impact of food safety standards on processed food exports from developing countries. Food Policy 34(5):445–457

Kapoor M, Ranjan P, Raychaudhuri J (2011) The impact of credit constraints on exporting firms: empirical evidence from India. Paper presented at the XXII annual conference on contemporary issues in development economics. Jadavpur University, Kolkata. SSRN 2023068

Kneller R, Pisu M (2007) Industrial linkages and export spillovers from FDI. The World Econ 30 (1):105–134

Konnings J (2001) The effects of foreign direct investment on domestic firms. Econ Transit 9 (3):619–633

Krishna P, Mitra D (1998) Trade liberalisation, market discipline and productivity growth: new evidence from India. J Dev Econ 56(2):447–462

Kumar N (1990) Multinational enterprises in India. Routledge, London

Kumar N (1994) Multinational enterprises and industrial organisation: a case of India. Sage, New Delhi

Kumar N, Siddharthan NS (1997) Technology, market structure and internationalisation: issues and policies for developing countries. Routledge, London

Kumar N, Aggarwal A (2005) Liberalization, outward orientation and in-house R&D activity of multinational and local firms: A quantitative exploration for indian manufacturing. Res Policy 34(4):441–460

Kumar N, Siddharthan NS (1994) Technology, firm size and export behaviour in developing countries: the case of Indian enterprises. J Dev Studies. 31(2):289–309

Kumar N, Pradhan JP (2003) Export competitiveness in knowledge based industries: a firm-level analysis of Indian manufacturing. *RIS Discussion Paper* No. 43. New Delhi: Research and Information System for Developing Countries

Manova K (2008) Credit constraints, heterogeneous firms, and international trade. NBER Working Paper No. 14531. National Bureau of Economic Research, Cambridge

Melitz MJ (2003) The impact of trade on Intra-industry reallocations and aggregate industry productivity. Econometrica 71(6):1695–1725

Mirabelle M (2008) Exporters and credit constraints: a firm-level approach. National Bank of Belgium Working Paper No. 139. Brussels

Pant M (1993) Export performance, transnational corporations and the manufacturing sector: a case study of India. Indian Econ Rev 28(1):41–54

Parameswaran M (2014) Productivity growth in a liberalising economy: evidence from indian manufacturing industry. In Kathuria V, Rajesh Raj SN, Sen K (eds) Productivity in Indian manufacturing-measurements, methods and analysis. Routledge, New Delhi

Paravisini D, Rappoport V, Schnabl P, Wolfenzen D (2011) Dissecting the effect of credit supply on trade: evidence from matched credit-export data. Working Paper 16975. National Bureau of Economic Research, Cambridge

Patibandala M (1995) Firm size and export behavior: an Indian case study. J Dev Studies 31 (6):868–882

Perez T (1997) Multinational enterprises and technological spillovers: an evolutionary model. J Evol Econ 7(2):169–192

Rasiah R (2003) How important is process and product technology capability in Malaysia and Thailand's FDI-Driven electronics industry? J Asian Econ 14(5):785–811

Rhee YW, Belot T (1990) Export catalysts in low income countries: a review of eleven success stories. World Bank Discussion Papers No. 72. Washington DC World Bank

Roper S, Love JH (2002) Innovation and export performance: evidence from the UK and German manufacturing plants. Res Policy 31(7):1087–1102

Ruane F, Sutherland J (2004) Ownership and export characteristics of irish manufacturing performance. IIIS Discussion Paper No. 32: Institute for International Integration Studies Trinity College Dublin, Dublin

Saggi K (2002) Trade, foreign direct investment and international technology transfer: a survey. World Bank Res Observer 17(2):197–235

Soete LLG (1981) A general test of technological gap trade theory. Rev World Econ 117(4):638–660

Sousa N, Greenaway D, Wakelin K (2000) Multinationals and export spillovers. University of Nottingham Research Paper 2000/14. Centre for Research on Globalisation and Labour Markets School of Economics University of Nottingham, Nottingham

Srinivasan TN, Archana V (2011) Determinants of export decision of firms. Econ Polit Wkly 56 (7):49–58

Swenson DL, Chen H (2014) Multinational exposure and the quality of new Chinese exports. Oxford Bull Econ Stat 76(1):41–66

Topalova P, Khandelwal A (2011) Trade liberalisation and firm productivity: the case of India. Rev Econ Stat 93(3):995–1009

Veugelers R, Cassiman B (1999) Make and buy in innovation strategies: evidence from Belgian manufacturing firms. Res Policy 28(1):63–80

Wang JY, Blomstrom M (1992) Foreign investment and technology transfer: a simple model. Eur Econ Rev 36(1):137–155

IPR Regulatory Policy, Commercial Piracy, and Entry Modes of MNC: A Theoretical Analysis

Nilanjana Biswas Mitra and Tanmoyee Banerjee Chatterjee

Abstract The present paper tries to link the IPR regulation rate chosen by an LDC government to the mode of entry chosen by an MNC firm in an LDC market where the options are export, fragmented production structure with assembly line FDI in LDC or complete production in LDC with full technology transfer. The last two options allow the possibility of entry of a pirate who can be deterred either by incurring anti-copying investment by the MNC or by the LDC government's enforcement in the form of monitoring. We show that if the transport cost of sending semifinished product from DC to LDC is below a critical level the fragmentation mode of entry is always chosen. Further if monitoring is socially optimal, entry of the pirate cannot be deterred.

Keywords Export · Assembly line FDI · FDI with complete production · IPR protection

JEL Classification L11 · O33 · O34 · O38

The paper is from the author's ongoing PhD dissertation at Jadavpur University.

N.B. Mitra
Department of Economics, Susil Kar College, 24 Prgs (South), Ghoshpur 743330, India

N.B. Mitra (✉) · T.B. Chatterjee
Department of Economics, Jadavpur University, Kolkata 700032, India
e-mail: bnilanjana2011@hotmail.com

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_4

1 Introduction

It is a common fact that the markets of developing economies are infested by cheap and identical replicas of popular products, thus eroding the profits of the original producers.[1] Hence an important consideration of the MNCs before initiation of production in such economies is the choice of mode of entry and adoption of suitable copy protection strategies. The role of IPR regulatory policy of the local government also plays a decisive role in making entry choice and technology transfer decisions. Thus our paper makes an effort to link the choice of mode of entry and copy protection strategy[2] adopted by a MNC to the IPR regulatory policy of the government.

The issue of entry choice and technology transfer have been dealt widely in the works of Vishwasrao (1994), Lin and Saggi (1999), Zigic (2000), Pack and Saggi (2001), Matto et al. (2004), Eicher and Kang (2005).

The paper by Vishwasrao (1994) formulates a model to discuss technology transfer in an environment where patent protection is uncertain. The paper incorporates asymmetric information in a screening game where the innovating firm can either export the product as a worldwide monopolist or license production to a southern firm alone as monopoly licensing. The result essentially depends on the patent protection of the southern government. The paper by Lin and Saggi (1999) has constructed a dynamic model in which at each instant two exporting firms (a leader and a follower) noncooperatively choose their probability of entry into a foreign market. Each firm generates cost-lowering externalities that help the other firm. These externalities thus generate an incentive to delay FDI. The paper by

[1]In case of software the study by Business Software Alliance shows in 2010 the worldwide piracy rate was 42 % as against the piracy rate of 64 % in a developing economy like India. Further the commercial value of the piracy in India is $2,739 million. In 2011 the global piracy rate was unchanged at 42 % while the commercial value of pirated PC software in India rose to $2.93 billion. http://globalstudy.bsa.org/2011/. Accessed on 24.5.2013.

[2]Empirical evidences show that in many occasions firm undertakes copy protection strategies by incurring anti-copying investment. Copy protection measures are available in a number of different types, for various forms of digital media, and are commonly referred to as Digital Rights Management Systems or DRMs. They describe a wide range of technical measures that are licensed for controlling, measuring and enabling use of copyright protected digital content for different purposes. Some systems provide levels of Rights Management Information (RMI), so that royalties can be collected on usage of the material depending on the context. Technical Protection Measures (TPM) are the most often used, as they are for copy protection of digital content. http://www.ehow.com/copy-protected-dvds/. Accessed on 14.5.2012.

In case of software, copy protection is implemented in several forms either on the physical media that comes with the media or within the software itself. Many programs try to control piracy through product keys which are a unique string of numbers that a user must enter to register the product http://www.ehow.com/list_7166954_software-copy-protection-techniques.html. Accessed on 14.5.2012.

Zigic (2000) analyzes the issue of optimal tariff when the northern and the southern firms compete in quantities in an imperfectly competitive northern market and there are potentially varying degrees of IPR violation by the south. The results show that three possible market structures can arise in equilibrium as a consequence of the tariff. Pack and Saggi (2001) provides a three-stage game to analyze the effect of vertical technology transfer on industrial development of LDCs with the possibility of technology diffusion. Matto et al. (2004) explore the preferences of a foreign firm and a welfare maximizing host country government over the two modes of FDI—direct entry or acquisition of existing domestic firms. This paper adopts a game-theoretic approach where in the first stage the foreign firm chooses the mode of entry. The result shows that the degree of technology transfer and the intensity of market competition depend upon the mode of entry chosen by the foreign firm. Eicher and Kang (2005) in their model examine how market structure, size, tariffs, and transport costs influence the entry behavior of MNCs. The model consists of three-stage entry game between a local firm and an MNC rival.[3]

However, none of the studies have dealt with fragmented production as one of the possible choice of mode of entry[4] and related the issue of optimal choice of entry mode of the MNC to the problem of commercial piracy which is the main focus of the present study. The present paper assumes that the MNC can enter the market of a developing country in three possible ways: they can export the finished product, can fragment the production process, and shift the assembly line works to the LDC[5] or can shift the entire production process to LDC through FDI. Given the options the paper tries to identify the optimal entry mode of a multinational firm in an LDC market, taking account of the fact that due to weak enforcement of IPR protection a pirate may exist in the economy. The model also endogenously determines IPR protection of the local LDC government and the copy protection investment level adopted by the MNC in response to the local government's IPR policy.

[3]Apart from these the issue of commercial piracy in a vertically differentiated market has been dealt widely in the works of Silve and Bernhardt (1998), Banerjee (2003), Poddar (2005), Banerjee et al. (2008), Kiema (2008), Banerjee (2011), Lu and Poddar (2012).

[4]Paper by Long (2005) explains why a parent company sets up a subsidiary in a low-wage economy and is outsourcing a part of its production thus, retaining core activities like design, patent applications and marketing in the original country ('Incomplete Outsourcing') in a vertically differentiated duopoly structure. However, here the issue of anti-copying investment is not addressed.

[5]Different empirical examples support the fact that in many cases the MNCs are shifting their assembly line works to LDCs with weak IPR framework. For example, recent news reveals the fact that multinational mobile phone company Nokia is shifting its assembly line works from Europe to Asia with plants in Europe concentrating on software-heavy aspects of the production process. http://articles.economictimes.indiatimes.com/2012-02-08/news/31037744_1_ceo-stephen-elop-nokia-cheaper-phones. http://www.indiatimes.com/mobile/nokia-to-shift-phone-assembly-to-asia-12119.html. Posted on 09 Feb 2012 AT 10:58:18. Accessed on 12.7.2012.

Our results show that in equilibrium either fragmented production structure with shifting of assembly line units to the LDC or FDI with full technology transfer can occur where the government may or may not monitor. However, monitoring may not result in the deterrence of piracy. Alternatively, under no-monitoring policy piracy can be deterred under FDI or fragmentation mode of entry when the MNC chooses to restrict copying completely. The paper specifies the conditions under which all these different equilibria hold.

The rest of the paper is organized as follows: Sect. 2 describes the model and the sequence of the game. Section 3 develops the model and equilibrium of the firm under different strategies. Section 4 gives the comparative analysis of different entry modes and optimal monitoring rate chosen by the local government. Section 5 presents a numerical analysis and Sect. 6 concludes the paper.

2 The Model

2.1 The Game Plan

The model considers an MNC having its origin in the Developed Country (DC) with the following options for production to initiate production in an LDC:

(1) It can produce entirely in the DC market and export the finished product to the LDC with a per unit positive transport cost.
(2) It can fragment the production process in two stages between the DC and the LDC. In the first stage, production of the core material takes place in the DC. In the second stage assembling of the core material takes place in the LDC.[6]
(3) It can undertake the entire production in the LDC by opening up the entire manufacturing and assembling unit with FDI.

In the second option the embodied technology and in the last option disembodied technology is transferred. In both the cases diffusion of technology results in the possibility of entry of the fake firm. The fake firm is capable of producing an identical replica of the original product. The probability of entry of the fake firm will, however, depend on the IPR protection exercised by the local government and the level of anti-copying investment undertaken by the MNC. The local LDC government as a monitoring authority will extract a penalty from the fake firm if detected. Now depending upon the government's IPR protection rate the MNC adopts two strategies, namely complete copy protection and accommodating strategy. In the paper, complete copy protection investment (CP strategy) is a

[6]A common example of this type of production is in the case of Coca Cola—one of world's leading beverage suppliers. The MNC prepares the concentrate in the United States which is then exported to different countries where the bottling units are located.

technology determined cost which prevents copying completely.[7] This is in contrast to anti-copying investment (AC strategy)[8] which makes the task of copying difficult.[9] In the former case the MNC takes an investment that completely prohibits the entry of fake firm and in the second case even with anti-copying investment fake firm enters the market and the game ends with the MNC acting as a price leader in a duopoly situation. The sequential game moves in the following manner. In the first stage of the game LDC government chooses IPR protection rate. Given this rate, MNC decides the mode of entry where the options are export, fragmented production, or FDI production. Under fragmentation and complete FDI production technology diffusion may cause the entry of fake firm in the market. Hence under these two cases MNC incurs a copy protection investment. Depending upon local government's IPR protection rate it adopts either a complete copy protection strategy or accommodating strategy. In case of accommodating strategy, the IPR protection rate and investment undertaken by the MNC determines the entry of fake firm in the market. If the fake firm enters, the MNC acts as a price leader in the duopoly set up. In case of complete copy protection strategy the MNC acts as a monopolist (Fig. 1).

Given the game tree, the game is solved using the backward induction method. At first the price game is solved. If the fake firm enters, MNC acts as the price leader and the fake firm follows, otherwise a monopoly price solution is obtained. Given the prices, the fake firm takes his entry decision which depends on the anti-copying investment of MNC and the monitoring rate exercised by the local government. Depending on the behavior of fake firm the MNC decides about its anti-copying investment for each entry mode. It compares its profit under the

---

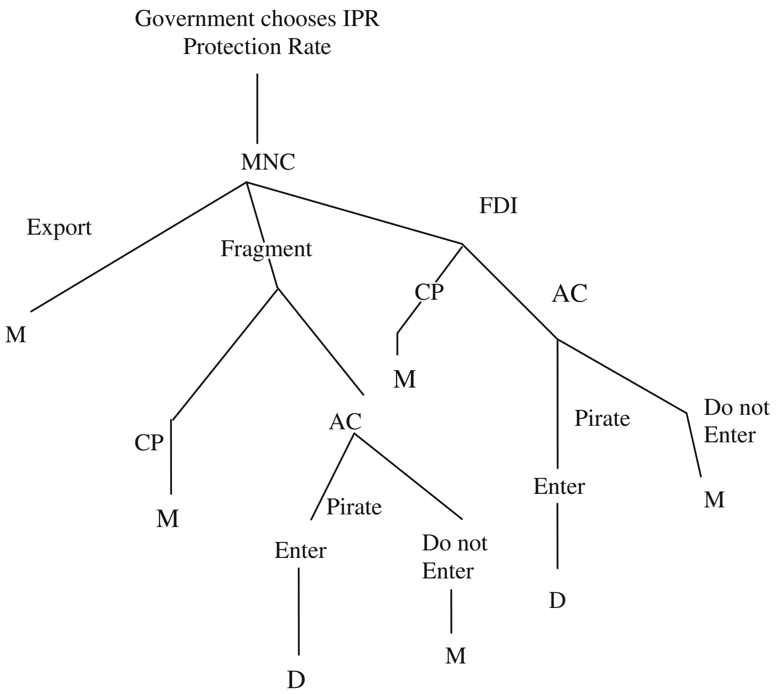
[7]Empirical evidences show that copy protection are available in a number of different types like digital licensing, digital watermarks, etc., for various forms of digital media, and are commonly referred to as Digital Rights Management Systems or DRMs. They describe a wide range of technical measures that are licensed for controlling, measuring, and enabling use of copyright protected digital content for different purposes. Technical Protection Measures (TPM) are the most often used, as they are for copy protection of digital content.

In case of software, copy protection is implemented in several forms either on the physical media that comes with the media or within the software itself. Many programs try to control piracy through product keys which are a unique string of numbers that a user must enter to register the product. Many game CDs and some audio CDs contain copy protection that prevents the disc from being copied or even played on other systems. http://www.ehow.com/how_6820532_detectcd-copy-protection.html, http://en.wikipedia.org/wiki/Digital_rights_management. Accessed on 09. 02.15.

[8]The AC strategy considers measures adopted by companies to make the task of copying harder or difficult. In this respect we must mention existence of 'Indian Music Industry' an association of all leading Indian recording companies. The companies have to pay a membership fee to become member of this association. This association undertakes anti-copying raids to deter piracy. http://indianmi.org/. Accessed on 09.2.2015.

Apart from this anti-copying investment may be done by encrypting commercial discs to make copying them a little more difficult. http://www.howtogeek.com/161498/. Accessed on 09.2.2015.

[9]On the basis of empirical evidences the paper considers CP and AC strategies are the two completely distinct strategies to combat piracy.



In all cases where pirate does not enter a monopoly outcome (M) is observed.
D denotes an outcome of price leadership game with the MNC as the price leader.

Fig. 1 The game tree

different strategies to decide the optimal entry mode which essentially depends on the monitoring rate chosen by the LDC government. Finally given the MNCs choice the LDC government chooses its optimal monitoring rate.

2.2 *The General Assumptions*

The model assumes that the product is sold and consumed solely in the LDC. In addition, the product under consideration is such that its production is fragmented in two parts. In the first part, the production of the core material takes place. In the second part, assembling or finishing of the core material takes place. The LDC government chooses a monitoring rate g, where cost of monitoring $C(g)$ is increasing and convex in g. Further it is assumed that complete monitoring by the government is costly $\underset{g\to 1}{Lt}\, C(g) \to \infty$.

The MNC is facing a linear demand function given as:

$$q = a - P. \tag{1}$$

where q is the quantity demanded, P is the price of the final product, and a is the market size parameter.

Given the general assumptions the three options for entry by MNC in LDC market are Export, Fragmentation, and FDI in LDC. The model assumes that for the last two entry modes the possibility of technology diffusion exists and a fake firm can enter the market. The model assumes that probability of copying of the original product by the fake firm is higher in FDI where complete transfer of technology takes place than in case of fragmentation where only embodied technology is transferred.

Given these assumptions the behavior of MNC and fake firm for different entry mode is analyzed in Sect. 3.

3 The Behavior of MNC and Fake Firm Under Different Entry Modes

3.1 Export Mode of Entry

First, we consider the situation where the MNC is undertaking the entire production process in her own country and exporting the final product to the LDC.

The total cost function of the MNC is defined as follows:

$$c_{\mathrm{DC}} = cq + A + tq$$

Here A be the sunk cost of production of core materials and c be the per unit cost of assembling or finishing the product. The finished product is then transferred to the LDC incurring a per unit transport cost t.

The profit of MNC is defined as follows:

$$\begin{aligned} \prod\nolimits_{\mathrm{Export}} &= pq - cq - tq - A \\ &= (a - q)q - cq - tq - A. \end{aligned} \tag{2}$$

From the First-Order profit maximizing conditions we get the monopoly output, price, and profit of the MNC as

$$q_{\mathrm{export}} = \frac{(a - c - t)}{2}, p_{\mathrm{export}} = \frac{(a + c + t)}{2}, \prod\nolimits_{\mathrm{export}} = \frac{(a - c - t)^2}{4} - A. \tag{3}$$

The Social Welfare:
SW = Consumer Surplus + Government Surplus[10]

$$\mathrm{SW}_{\mathrm{export}} = \frac{(a-c-t)^2}{8} - C(g). \tag{4}$$

3.2 Fragmentation Mode of Entry

In this case MNC chooses to fragment the production in such a way that it conducts the manufacturing of core parts in its own country (thus bringing in embodied technology to the LDC) and complete the assembling part in the LDC. As mentioned earlier the MNC can adopt two alternative strategies namely:

Complete Copy Protection strategy—The MNC undertakes anti-copying investment such that entry of the fake producer is prevented ensuring monopoly profit for the MNC. We assume that $x_{\mathrm{frag}}^{\mathrm{cp}}$ be the level of anti-copying investment which completely prohibits the illegal copying of the MNCs product and thus rules out the entry of fake firm in the LDC market.

The total cost function under Complete Copy Protection (CP) strategy is

$$c_{\mathrm{frag}}^{\mathrm{cp}} = wq + A + tq + x_{\mathrm{frag}}^{\mathrm{cp}}. \tag{5}$$

where w is the per unit cost of assembling the semifinished product in the LDC[11] and t is the per unit transport cost to transfer the intermediate product to the LDC. For simplicity it is assumed to be same as the transport cost of the finished product when the MNC is simply exporting the product.

The profit of the MNC in this case will be

$$\begin{aligned} \prod\nolimits_{\mathrm{frag}}^{\mathrm{cp}} &= \prod\nolimits_{\mathrm{frag}}^{\mathrm{cp}(m)} - \bar{x}_{\mathrm{frag}}^{\mathrm{cp}} \\ &= \left[p_{\mathrm{frag}}^{\mathrm{cp}} - (w+t)\right] q_{\mathrm{frag}}^{\mathrm{cp}} - A - \bar{x}_{\mathrm{frag}}^{\mathrm{cp}}. \end{aligned} \tag{6}$$

The monopoly profit, price, and quantity will be

$$\prod\nolimits_{\mathrm{frag}}^{\mathrm{cp}} = \frac{\{a-(w+t)\}^2}{4} - A - \bar{x}_{\mathrm{frag}}^{\mathrm{cp}}, q_{\mathrm{frag}}^{\mathrm{cp}} = \frac{a-w-t}{2}, p_{\mathrm{frag}}^{\mathrm{cp}} = \frac{a+w+t}{2}. \tag{7}$$

---

[10]Where Social Welfare is maximized at $g = 0$.

[11]It is assumed that $w < c$ due to cheap labor in the LDC.

The social welfare in this case will be

$$\begin{aligned} \mathrm{SW}_{\mathrm{frag}}^{\exp(\mathrm{cp})} &= \text{Consumer Surplus} + \text{Government Surplus} \\ &= \frac{\{a-(w+t)\}^2}{8} - C(g). \end{aligned} \tag{8}$$

Accommodating Strategy—In this case anti-copying investment $x_{\mathrm{frag}}^{\mathrm{ac}}$ is undertaken by the MNC in such a way that the fake producer can enter and operate in the market. The MNC acts as a price leader with the fake producer acting as a follower. Let $\mathrm{kh}(x_{\mathrm{frag}}^{\mathrm{ac}})$ be the probability of copying the original product where $0<k<1$; $0<h(x)<1$; $h'(x)<0$; $h''(x) > 0 : 0<kh(x)<1$; $h(x) = 1 \text{ for } x = 0$; $\underset{x\to\infty}{Lt}\, h(x_{\mathrm{frag}}^{\mathrm{ac}}) \to 0$ and k is the exogenously given copying parameter.[12] Thus in the accommodating strategy complete deterrence of the pirate is ruled out. The model also makes the following assumption:

$$\mathrm{A1} : h(x)h''(x) - \{h'(x)\}^2 > 0$$

The probability that the fake producer will be detected is g where g is the local government's monitoring rate of IPR. Therefore, the probability of entry of the fake producer can be written as $kh(x_{\mathrm{frag}}^{\mathrm{ac}})(1-g)$ and the probability of detection of the fake producer is $\left\{1-kh(x_{\mathrm{frag}}^{\mathrm{ac}})(1-g)\right\}$.

The total cost function of the MNC is given by

$$c_{\mathrm{frag}}^{\mathrm{ac}} = wq + A + tq + x_{\mathrm{frag}}^{\mathrm{ac}}$$

The profit of the MNC when the fake producer enters and operates in the market

$$\prod\nolimits_{\mathrm{frag}}^{\mathrm{ac}} = \left[p_{\mathrm{frag}}^{\mathrm{ac}} - (w+t)\right] q_{\mathrm{frag}}^{\mathrm{ac}} - A - x_{\mathrm{frag}}^{\mathrm{ac}}. \tag{9}$$

For the fake producer profit is defined as:

$$\prod\nolimits_{\mathrm{frag}}^{\mathrm{fake}} = \left[p_{\mathrm{frag}}^{\mathrm{ac}} q_{\mathrm{frag}}^{\mathrm{fake}} - \frac{1}{2} q_{\mathrm{frag}}^{2\,\mathrm{fake}}\right] - F. \tag{10}$$

where the cost functions of the fake firm is

$$C(q_{\mathrm{frag}}^{\mathrm{fake}}) = \frac{q_{\mathrm{frag}}^{2\mathrm{fake}}}{2} + F. \tag{10a}$$

and F be the sunk cost.

[12] 'k' is assumed to be a fraction because under fragmentation mode of entry as full technology transfer is not taking place a firm cannot copy the product with certainty if the firm is not investing in anti-copying investment.

Solving the price leadership game by backward induction method we first determine the FOC of profit maximization of the fake producer and equate it to zero

$$\frac{\partial \pi_{\text{frag}}^{\text{fake}}}{\partial q_{\text{frag}}^{\text{fake}}} = 0 \Rightarrow p_{\text{frag}}^{\text{ac}} = q_{\text{frag}}^{\text{fake}}. \tag{11}$$

The residual demand function for the MNC will be

$$q_{\text{frag}}^{\text{ac}} = a - p_{\text{frag}}^{\text{ac}} - q_{\text{frag}}^{\text{fake}}. \tag{12}$$

Replacing Eq. (11) in (9) and solving for the profit maximizing equilibrium values we get[13]

$$\begin{aligned} &\prod\nolimits_{\text{frag}}^{\text{ac}} = \frac{\{a-2(w+t)\}^2}{8} - A - x_{\text{frag}}^{\text{ac}}, \quad \prod\nolimits_{\text{frag}}^{\text{fake}} = \frac{\{a+2(w+t)\}^2}{32} - F \\ &q_{\text{frag}}^{\text{ac}} = \frac{(a-2(w+t))}{2}\, q_{\text{frag}}^{\text{fake}} = \frac{a+2(w+t)}{4} = p_{\text{frag}}^{\text{ac}}. \end{aligned} \tag{13}$$

In case the fake producer gets detected due to IPR protection exercised by the local government then the profit equation of the MNC will be

$$\begin{aligned} &\prod\nolimits_{\text{frag}}^{\text{ac}(m)} = \left[p_{\text{frag}}^{\text{ac}(m)} - (w+t)\right] q_{\text{frag}}^{\text{ac}(m)} - A - x_{\text{frag}}^{\text{ac}} \\ &\prod\nolimits_{\text{frag}}^{\text{ac}(m)} = \frac{\{a-(w+t)\}^2}{4} - A - x_{\text{frag}}^{\text{ac}}, \; q_{\text{frag}}^{\text{ac}(m)} = \frac{a-w-t}{2}, \; p_{\text{frag}}^{\text{ac}(m)} = \frac{a+w+t}{2}. \end{aligned} \tag{14}$$

From the two profit functions for the MNC given in Eqs. (13) and (14) the expected profit function for the MNC can be given as

$$\begin{aligned} \prod\nolimits_{\text{frag}}^{\text{ac(exp)}} &= kh\left(x_{\text{frag}}^{\text{ac}}\right)(1-g)\prod\nolimits_{\text{frag}}^{\text{ac}} + \left\{1 - kh(x_{\text{frag}}^{\text{ac}})(1-g)\right\}\prod\nolimits_{\text{frag}}^{\text{ac}(m)} \\ &= \frac{\{a-(w+t)\}^2}{4} - kh\left(x_{\text{frag}}^{\text{ac}}\right)(1-g)\left\{\frac{\{a-(w+t)\}^2}{4} - \frac{\{a-2(w+t)\}^2}{8}\right\} - A - x_{\text{frag}}^{\text{ac}}. \end{aligned} \tag{15}$$

Maximizing (15) MNC determines the optimal value of anti-copying investment ($\bar{x}_{\text{frag}}^{\text{ac}}$) which satisfies (16).[14]

[13]The duopoly profit of MNC under fragmentation with AC strategy will be positive if and only if $a > 2(w + t)$.

[14]Second-order condition requires that $h''(x) > 0$.

$$-kh'(\bar{x}^{ac}_{frag})(1-g)\left[\frac{\{a-(w+t)\}^2}{4}-\frac{\{a-2(w+t)\}^2}{8}\right]=1. \tag{16}$$

Proposition 1 *In the Fragmentation mode of entry, under Accommodating Strategy the anti-copying investment undertaken by the MNC ($\bar{x}^{ac}_{frag}$) is inversely related to IPR protection rate (g) and transport cost (t).*

Proof See Appendix 1.

Thus if the local LDC government's provision of IPR protection or the transport cost of transferring the semifinished good from the DC to the LDC increases then the MNC will reduce its anti-copying investment in case of AC strategy. There is a substitutability between local government's provisions of IPR protection vis-á-vis that of MNC firm. Alternatively, a rise in t increases the marginal cost of assembling the product for LDC. Thus profitability requires a cut in anti-copying investment by the MNC.

The expected profit function for the fake firm will be

$$\prod\nolimits_{frag}^{fake(exp)}=kh\left(x^{ac}_{frag}\right)(1-g)\prod\nolimits_{frag}^{fake}+\left\{1-h\left(x^{ac}_{frag}\right)\right\}\cdot 0-kh\left(x^{ac}_{frag}\right)\cdot gG-F$$

where G is the lump sum penalty which the fake firm has to pay the local LDC government if it gets detected.

Replacing the values from Eq. (13) we get

$$\prod\nolimits_{frag}^{fake(exp)}=kh\left(x^{ac}_{frag}\right)\left[(1-g)\frac{\{a+2(w+t)\}^2}{32}-gG\right]-F. \tag{17}$$

Proposition 2

(i) *Fake firm will not operate under Fragmentation for $g\in[g^*_{frag},1]$ where g^*_{frag} is defined in Eq. (18).*

(ii) g^*_{frag} *increases with t and decreases with penalty level G where*

$$g^*_{frag}=\frac{\{a+2(t+w)\}^2-\frac{32F}{kh\left(x^{ac}_{frag}\right)}}{\{a+2(w+t)\}^2+32G}. \tag{18}$$

(iii) *Profit of the fake producer decreases unambiguously with the IPR monitoring rate if the assumption A1 holds.*[15]

Proof See Appendix 1.

That is if $g \geq g^*_{\text{frag}}$ then fake firm does not enter. In other words for a sufficiently high value of g the MNC enjoys monopoly profit and also does not have to undertake any anti-copying investment. Second, a high penalty lowers the profitability and deters the entry of fake firm. Alternatively as t increases, per unit transport cost for MNC increases, raising the price. A higher price increases the profitability of fake producer. Hence g^*_{frag} rises, implying that a high monitoring rate is required to deter entry of fake producer.

The profit of the fake producer decreases with 'g' the IPR monitoring rate if the assumption A1 holds. A rise in g reduces the expected profit of the fake firm at an unchanged level of anti-copying investment. However, from Proposition 1 we know that a rise in g reduces the anti-copying investment thereby increasing the fake firm's profit. But the overall effect will be negative if condition A1 holds.

The social welfare[16] will be

$$\text{SW}^{\text{ac(exp)}}_{\text{frag}} = \text{Consumer Surplus} + \prod\nolimits^{\text{fake}}_{\text{frag}} + \text{Government Surplus}$$

Now Consumer Surplus is given as

$$\text{CS}^{\exp(\text{ac})}_{\text{frag}} = (1-g)kh\left(x^{\text{ac}}_{\text{frag}}\right)\frac{\{3a-2(w+t)\}^2}{32} + \frac{\left\{1-kh(x^{\text{ac}}_{\text{frag}})(1-g)\right\}}{8}\{a-(w+t)\}^2 - F$$

The Government Surplus can be given as

$$\text{Government Surplus} = kh(x^{\text{ac}}_{\text{frag}})gG - C(g)$$

where G is the local LDC government's earning received when the fake firm gets detected and $C(g)$ is the cost of providing IPR protection.

Summing up Consumer Surplus, profit of the fake firm (which retains its entire profit in the LDC) and Government Surplus we get the Social Welfare of the AC strategy as

[15]A1 rules out the situation where a stringent IPR monitoring rate improves the profit of the fake firm.

[16]It may be argued that since fake firm is not socially desirable so its profit should be excluded from Social Welfare calculations. However, since the fake firm generates a surplus in the host country so its profit is included in the above calculations. Even if we exclude the fake firm from social welfare calculations, the results remain essentially the same.

$$
\begin{aligned}
\mathrm{SW}^{\mathrm{ac(exp)}}_{\mathrm{frag}} &= \frac{(a-w-t)^2}{8} + (1-g)kh(x^{\mathrm{ac}}_{\mathrm{frag}})\left[\frac{(3a-2(w+t))^2}{32} + \frac{(a+2(w+t))^2}{32} - \frac{(a-w-t)^2}{8}\right] \\
&\quad - C(g) - F \\
&= \frac{(a-w-t)^2}{8} + (1-g)kh(x^{\mathrm{ac}}_{\mathrm{frag}})\Gamma - C(g) - F \\
\text{where } \Gamma &= \left[\frac{(3a-2(w+t))^2}{32} + \frac{(a+2(w+t))^2}{32} - \frac{(a-w-t)^2}{8}\right] \\
&= \frac{6a^2+4(w+t)^2}{32} > 0.
\end{aligned}
\tag{19}
$$

$$
\frac{\partial \mathrm{SW}^{\mathrm{ac(exp)}}_{\mathrm{frag}}}{\partial g} = -\frac{k\Gamma}{h''(x^{\mathrm{ac}}_{\mathrm{frag}})}\left\{h(x^{\mathrm{ac}}_{\mathrm{frag}})h''(x^{\mathrm{ac}}_{\mathrm{frag}}) - \left\{h'(x^{\mathrm{ac}}_{\mathrm{frag}})\right\}^2\right\} - C'(g) < 0. \tag{19a}
$$

if assumption A1 holds.

Comparative Study of Copy Protection and Accommodating Strategies under Fragmentation Mode of Entry

Comparing (7) and (15) we find

$$
\begin{aligned}
\prod\nolimits^{\mathrm{cp}}_{\mathrm{frag}} - \prod\nolimits^{\mathrm{ac(exp)}}_{\mathrm{frag}} &= kh\left(x^{\mathrm{ac}}_{\mathrm{frag}}\right)(1-g)\left\{\frac{(a-w-t)^2}{4} - \frac{\{a-2(w+t)\}^2}{8}\right\} \\
&\quad - \left(\bar{x}^{\mathrm{cp}}_{\mathrm{frag}} - \bar{x}^{\mathrm{ac}}_{\mathrm{frag}}\right).
\end{aligned}
\tag{20}
$$

Equation (20) shows that if equilibrium value of $\bar{x}^{\mathrm{ac}}_{\mathrm{frag}}$ is greater than or equal to $\bar{x}^{\mathrm{cp}}_{\mathrm{frag}}$ the AC strategy is always dominated by the CP strategy.

Again differentiating (20) with respect to 'g'

$$
\frac{\partial(\prod^{\mathrm{cp}}_{\mathrm{frag}} - \prod^{\mathrm{ac(exp)}}_{\mathrm{frag}})}{\partial g} = -kh(x^{\mathrm{ac}}_{\mathrm{frag}})\left\{\frac{(a-w-t)^2}{4} - \frac{\{a-2(w+t)\}^2}{8}\right\} < 0. \tag{21}
$$

The above result shows that AC strategy in the fragmented mode of Entry becomes more profitable as 'g' goes up. Thus for $\bar{x}^{\mathrm{ac}}_{\mathrm{frag}} < \bar{x}^{\mathrm{cp}}_{\mathrm{frag}}$ let us assume that there exists a value of g say $\tilde{g}_{\mathrm{frag}}$ such that $\prod^{\mathrm{cp}}_{\mathrm{frag}} = \prod^{\mathrm{ac(exp)}}_{\mathrm{frag}}$. Solving for that value of g we get the following results.

$$
\tilde{g}_{\mathrm{frag}} = 1 - \frac{4(\bar{x}^{\mathrm{cp}}_{\mathrm{frag}} - \bar{x}^{\mathrm{ac}}_{\mathrm{frag}})}{kh(\bar{x}^{\mathrm{ac}}_{\mathrm{frag}})\frac{a^2-2(w+t)^2}{2}} = 1 - \frac{8(\bar{x}^{\mathrm{cp}}_{\mathrm{frag}} - \bar{x}^{\mathrm{ac}}_{\mathrm{frag}})}{kh(\bar{x}^{\mathrm{ac}}_{\mathrm{frag}})(a^2-2(w+t)^2)} \tag{22}
$$

Therefore depending on the value of 'g' we get the following strategy choice for the MNC when it considers the fragmented mode of entry.[17] Otherwise complete copy protection strategy will be a dominated strategy for all values of g. Table 1 summarizes the result.

Since $\frac{\partial \mathrm{SW}_{\mathrm{frag}}^{\mathrm{ac(exp)}}}{\partial g} < 0$ the social welfare under AC strategy is maximized at $g = \tilde{g}_{\mathrm{frag}}$.

3.3 Complete Production in LDC

In this option the MNC undertakes complete production process in LDC with full technology transfer, thus the probability of entry of the fake producer is much higher in this option. Hence the anti-copying investment incurred by the MNC to deter the entry of the fake producer will also be higher. But the MNC does not have to incur any transport cost either to transport the finished or semifinished product to the LDC.

As in case of Fragmented Production, in this mode of entry also the MNC will have two alternative strategies to choose from, namely Accommodating (AC) Strategy and Complete Copy protection (CP) Strategy.

Complete Copy Protection Strategy: The MNC incurs $x_{\mathrm{LDC}}^{\mathrm{cp}}$ level of anti-copying investment to completely prohibit the illegal copying of the MNCs product thus ruling out the entry of fake firm in the LDC market. We assume $x_{\mathrm{LDC}}^{\mathrm{cp}} > x_{\mathrm{frag}}^{\mathrm{cp}}$ since in this strategy copying probability is higher than that of the Fragmented mode of entry.

The profit of the MNC in this case will be

$$\prod\nolimits_{\mathrm{LDC}}^{\mathrm{cp}} = \prod\nolimits_{\mathrm{LDC}}^{\mathrm{cp(m)}} - x_{\mathrm{LDC}}^{\mathrm{cp}} = \left[p_{\mathrm{LDC}}^{\mathrm{cp}} - w\right] q_{\mathrm{LDC}}^{\mathrm{cp}} - A - x_{\mathrm{LDC}}^{\mathrm{cp}}. \tag{23}$$

The monopoly profit, price, and quantity will be

$$\prod\nolimits_{\mathrm{LDC}}^{\mathrm{cp}} = \frac{\{a-w\}^2}{4} - A - \bar{x}_{\mathrm{LDC}}^{\mathrm{cp}}, \quad q_{\mathrm{LDC}}^{\mathrm{cp}} = \frac{a-w}{2}, \quad p_{\mathrm{LDC}}^{\mathrm{cp}} = \frac{a+w}{2}. \tag{24}$$

The social welfare in this case will be

$$\mathrm{SW}_{\mathrm{LDC}}^{\mathrm{cp}} = \text{Consumer Surplus} = \frac{\{a-w\}^2}{8} - C(g). \tag{25}$$

---

[17] In this case we have implicitly assumed that $\bar{x}_{\mathrm{cp}}$ is not very high so that $\prod\nolimits_{\mathrm{frag}}^{\mathrm{cp}} = \frac{\{a-(w+t)\}^2}{4} - A >$
$$= kh(x_{\mathrm{frag}}^{\mathrm{ac}}(g=0))\frac{\{a-2(w+t)\}^2}{8} + \left\{1 - kh(x_{\mathrm{frag}}^{\mathrm{ac}}(g=0))\right\}\frac{(a-w-t)^2}{4} - x_{\mathrm{frag}}^{\mathrm{ac}}(g=0) - A = \prod\nolimits_{\mathrm{frag}}^{\mathrm{ac(exp)}}(g=0).$$

Table 1 Summary of results for fragmentation mode of entry

IPR monitoring rate	MNC's profit	Social welfare
$g \geq g^{*}_{\text{frag}}$ where $g^{*}_{\text{frag}} = \frac{\{a+2(w+t)\}^2 - \frac{32F}{kh(x^{\text{ac}}_{\text{frag}})}}{\{a+2(w+t)\}^2 + 32G}$	The fake producer does not enter thereby resulting in monopoly profit for the MNC $\prod^{\text{cp}}_{\text{frag}} = \frac{\{a-(w+t)\}^2}{4} - A$	$SW^{\text{mono}}_{\text{frag}} = \frac{\{a-w-t\}^2}{8} - C(g)$
$\tilde{g}_{\text{frag}} \leq g < g^{*}_{\text{frag}}$ where $\tilde{g}_{\text{frag}} = 1 - \frac{8(\bar{x}^{\text{cp}}_{\text{frag}} - \bar{x}^{\text{ac}}_{\text{frag}})}{kh(\bar{x}^{\text{ac}}_{\text{frag}})(a^2 - 2(w+t)^2)}$	The MNC chooses AC Strategy $\prod^{\text{ac(exp)}}_{\text{frag}} = \frac{\{a-(w+t)\}^2}{4} - kh(x^{\text{ac}}_{\text{frag}})(1-g)\left\{\frac{\{a-(w+t)\}^2}{4} - \frac{\{a-2(w+t)\}^2}{8}\right\} - A - x^{\text{ac}}_{\text{frag}}$	$SW^{\text{ac(exp)}}_{\text{frag}} = \frac{(a-w-t)^2}{8} + (1-g)kh(\bar{x}^{\text{ac}}_{\text{frag}})\Gamma - C(g) - F$ where $\Gamma = \frac{6a^2 + 4(w+t)^2}{32}$
$g < \tilde{g}_{\text{frag}}$	The MNC adopts CP strategy $\prod^{\text{cp}}_{\text{frag}} = \frac{\{a-(w+t)\}^2}{4} - A - \bar{x}^{\text{cp}}_{\text{frag}}$	$SW^{\text{cp}}_{\text{frag}} = \frac{\{a-(w+t)\}^2}{8}$

Accommodating Strategy—It is assumed that $h(x^{ac}_{LDC})$ is the probability of copying the original product where x^{ac}_{LDC} is the anti-copying investment made by the MNC such that $h'(.) < 0$ and $h''(.) > 0$[18] and $\underset{x\to\infty}{Lt}\ h(x^{ac}_{LDC}) \to 0$.

Like the fragmentation mode here also the pirate incurs a sunk cost F and has to pay G the cost/penalty to the LDC government in case it gets detected.

In this strategy, if fake firm enters and operates then the MNC acting as a price leader sets the price and the fake producer acts a follower. Solving for the profit maximizing expressions price, profit are defined in (26).

$$\prod\nolimits^{ac}_{LDC} = \frac{(a-2w)^2}{8} - A - x^{ac}_{LDC,} \quad q^{ac}_{LDC} = \frac{a-2w}{2}, \quad q^{fake}_{LDC} = p^{LDC}_{ac} = \frac{a+2w}{4}. \tag{26}$$

In case when the fake firm after entering is detected by the IPR protection policy of the local government then the profit maximizing expressions are

$$\prod\nolimits^{ac(m)}_{LDC} = \frac{(a-w)^2}{4} - A - x^{ac}_{LDC}, \quad q^{ac(m)}_{LDC} = \frac{a-w}{2}, \quad p^{ac(m)}_{LDC} = \frac{a+w}{2}. \tag{27}$$

The expected profit of the MNC is therefore given by

$$\prod_{LDC}^{ac(exp)} = \frac{\{a-w\}^2}{4} - h(\bar{x}^{ac}_{LDC})(1-g)\left[\frac{\{a-w\}^2}{4} - \frac{\{a-2w\}^2}{4}\right] - A - x^{ac}_{LDC}. \tag{28}$$

Maximizing (28) with respect to x^{ac}_{LDC} solves for optimum level of anti-copying investment $\bar{x}^{ac}_{LDC}$, that is, given by (29),

$$-h'(\bar{x}^{ac}_{LDC})(1-g)\left[\frac{(a-w)^2}{4} - \frac{(a-2w)^2}{8}\right] = 1. \tag{29}$$

Proposition 3 *Under FDI, with Accommodating Strategy the anti-copying investment undertaken by the MNC ($\bar{x}^{ac}_{LDC}$) will be inversely related to IPR protection (g).*[19]

The profit of fake produce is given by (30).

$$\prod\nolimits^{fake}_{LDC} = h(x^{ac}_{LDC})\left[(1-g)\frac{(a+2w)^2}{32} - gG\right] - F. \tag{30}$$

---

[18]Since, $0 < k < 1$ therefore for same level of 'x' the probability of copying in the FDI Production is more than that of the Fragmented mode of entry.

[19]Proof is same as that of Proposition 1, where $t = 0$ and $k = 1$.

Proposition 4[20]

(i) *Fake firm will not operate under FDI mode of entry for* $g \in [g^{*}_{\mathrm{LDC}}, 1]$
(ii) g^{*}_{LDC} *decreases with penalty level G.*
where $g^{*}_{\mathrm{LDC}} = \frac{\{a+2w\}^2 - 32F/h(x^{ac}_{\mathrm{LDC}})}{\{a+2w)\}^2 + 32G}$
(iii) *Profit of the fake producer decreases unambiguously with the monitoring rate if the assumption A1 holds.*

Expected social welfare given in Eq. (31) is as follows:

$$\begin{aligned} \mathrm{SW}^{ac(exp)}_{\mathrm{LDC}} &= \frac{(a-w)^2}{8} + (1-g)h(x^{ac}_{\mathrm{LDC}})\left[\frac{(3a-2w)^2}{32} + \frac{(a+2w)^2}{32} - \frac{(a-w)^2}{8}\right] - C(g) - F \\ &= \frac{(a-w)^2}{8} + (1-g)h(x^{ac}_{\mathrm{LDC}})T - C(g) - F. \end{aligned} \tag{31}$$

$$\text{where} \quad T = \left[\frac{(3a-2w)^2}{32} + \frac{(a+2w)^2}{32} - \frac{(a-w)^2}{8}\right] = \frac{6a^2 + 4w^2}{32} > 0$$

$$\frac{\partial \mathrm{SW}^{ac(exp)}_{\mathrm{LDC}}}{\partial g} = -\frac{T}{h''(^{ac}_{\mathrm{LDC}})}\left\{h(x^{ac}_{\mathrm{LDC}})h''(x^{ac}_{\mathrm{LDC}}) - \left\{h'(x^{ac}_{\mathrm{LDC}})\right\}^2\right\} - C'(g) < 0$$

if assumption A1 holds.

Comparative Study of Copy Protection and Accommodating Strategies under FDI production

Comparing the two profit equations Eqs. (24) and (28) under the AC and CP strategies we find that AC strategy becomes more profitable as 'g' goes up. Again solving for the value of 'g' for which $\prod^{cp}_{\mathrm{LDC}} = \prod^{ac(exp)}_{\mathrm{LDC}}$ we get

$$\tilde{g}_{\mathrm{LDC}} = 1 - \frac{8(\bar{x}^{cp}_{\mathrm{LDC}} - \bar{x}^{ac}_{\mathrm{LDC}})}{h(\bar{x}^{ac}_{\mathrm{LDC}})(a^2 - 2w^2)}. \tag{32}$$

Table 2 summarizes the results for the Complete production in the LDC:

Since $\frac{\partial \mathrm{SW}^{ac(exp)}_{\mathrm{LDC}}}{\partial g} < 0$ the social welfare under AC strategy is maximized at $g = \tilde{g}_{\mathrm{LDC}}$.

[20]Proof is same as that of Proposition 2, where $t = 0$ and $k = 1$.

Table 2 Summary of results for complete ldc mode of entry

IPR monitoring rate	MNC's profit	Social welfare
$g > g^*_{\mathrm{LDC}}$ where $g^*_{\mathrm{LDC}} = \frac{\{a+2w\}^2 - 32F/h(x^{\mathrm{ac}}_{\mathrm{LDC}})}{\{a+2w)\}^2 + 32G}$	The fake producer does not enter thereby resulting in monopoly profit for the MNC. $\prod^{\mathrm{mono}}_{\mathrm{LDC}} = \frac{\{a-w)\}^2}{4} - A$	$\mathrm{SW}^{\mathrm{mono}}_{\mathrm{LDC}} = \frac{\{a-w\}^2}{8} - C(g)$
$\tilde{g}_{\mathrm{LDC}} \leq g \leq g^*_{\mathrm{LDC}}$ where $\tilde{g}_{\mathrm{LDC}} = 1 - \frac{8(\bar{x}^{\mathrm{cp}}_{\mathrm{LDC}} - \bar{x}^{\mathrm{ac}}_{\mathrm{LDC}})}{h(\bar{x}^{\mathrm{ac}}_{\mathrm{LDC}})(a^2 - 2w^2)}$	The MNC chooses AC Strategy $\prod^{\mathrm{ac(exp)}}_{\mathrm{LDC}} = \frac{\{a-w\}^2}{4} - h(\bar{x}^{\mathrm{ac}}_{\mathrm{LDC}})(1-g)\left[\frac{\{a-w\}^2}{4} - \frac{\{a-2w\}^2}{4}\right] - A - x^{\mathrm{ac}}_{\mathrm{LDC}}$	$\mathrm{SW}^{\mathrm{ac(exp)}}_{\mathrm{LDC}} \frac{(a-w)^2}{8} + (1-g)h((\bar{x}^{\mathrm{ac}}_{\mathrm{LDC}})T - C(g) - F$ where $T = \frac{6a^2 + 4w^2}{32}$
$g < \tilde{g}_{\mathrm{LDC}}$	The MNC adopts CP strategy $\prod^{\mathrm{cp}}_{\mathrm{LDC}} = \frac{\{a-w\}^2}{4} - A - x^{\mathrm{LDC}}_{\mathrm{cp}}$	$\mathrm{SW}^{\mathrm{cp}}_{\mathrm{LDC}} = \frac{\{a-w\}^2}{8}$

4 Choice of Optimal Entry Modes and IPR Policy

Here we present an analysis to determine the sub-game perfect Nash equilibrium solution of the model. For simplicity of analysis we assume that export mode of entry is a dominated strategy compared to the other two modes of entry for high value of 'c', where 'c' be per unit variable cost of finishing the product in the DC. Hence subsequent analysis considers only Fragmentation and FDI modes of entry. Let g_E be the optimal monitoring rate chosen by the LDC government.

To determine the equilibrium solution we define a critical value of transport cost 't' given as $\tilde{t}$ in Eq. (33) such that at $t = \tilde{t}$, profit of the MNC under the CP Strategy for Fragmentation and FDI modes of entry are equal, that is, $\Pi^{\text{cp}}_{\text{frag}} = \Pi^{\text{cp}}_{\text{LDC}}$.

$$\tilde{t} = (a - w) - \sqrt{(a - w)^2 - 4(\bar{x}^{\text{cp}}_{\text{LDC}} - \bar{x}^{\text{cp}}_{\text{frag}})}. \tag{33}$$

Thus, for $t < \tilde{t}$ we have $\Pi^{\text{cp}}_{\text{frag}} = \Pi^{\text{cp}}_{\text{LDC}}$ and vice versa.

We define a locus (KT) of 'k' and 't' combinations along which $\prod^{\text{ac(exp)}}_{\text{frag}}\Big|_{g=\tilde{g}_{\text{LDC}}} = \prod^{\text{cp}}_{\text{LDC}}$, where the slope of the locus as defined below[21]:

$$\left.\frac{\mathrm{d}k}{\mathrm{d}t}\right|_{KT} = -\frac{\partial \prod^{\text{ac(exp)}}_{\text{frag}} / \partial t}{\partial \prod^{\text{ac(exp)}}_{\text{frag}} / \partial k}$$

By differentiating (15) with respect to 't' and 'k' we have $\frac{\partial \Pi^{\text{ac(exp)}}_{\text{frag}}}{\partial t} < 0$[22] and $\frac{\partial \Pi^{\text{ac(exp)}}_{\text{frag}}}{\partial k} < 0$.

Thus $\left.\frac{\mathrm{d}k}{\mathrm{d}t}\right|_{KT} < 0$.

For 'k' and 't' combinations lying above the locus we have $\prod^{\text{ac(exp)}}_{\text{frag}}\Big|_{g=\tilde{g}_{\text{LDC}}} < \prod^{\text{cp}}_{\text{LDC}}$ as high 'k' and 't' reduce the Fragmented profit of the MNC under the AC strategy and vice versa.

Proposition 5 *There exists a value of shipment cost 't' defined by $\tilde{t}$ in Eq. (33), such that*

For $t \leq \tilde{t}$, Fragmented mode of entry will always be chosen in equilibrium.

(i) *For $t > \tilde{t}$, and t and k combinations lying above the KT locus $g_E = \tilde{g}_{\text{LDC}}$ and FDI production with accommodating strategy will be chosen in equilibrium.*

---

[21]KT schedule will cut t axis for $k = 0$ at $t_1* = (a - w) - \sqrt{(a - w)^2 - 4\bar{x}^{\text{cp}}_{\text{LDC}}}$.

[22]If 'a' the market size parameter is sufficiently large then the condition holds unambiguously.

(ii) *For $t > \tilde{t}$ and t and k combinations lying below the KT locus, $g_E = 0$ and FDI production with complete copy protection strategy will be chosen in equilibrium.*

Proof

(i) $t \leq \tilde{t} \Rightarrow \prod^{cp}_{\mathrm{LDC}} \leq \prod^{cp}_{\mathrm{frag}}$. This implies that profit under the Fragmentation mode of entry dominates that of the FDI mode.[23]

Under this configuration, if $\mathrm{SW}^{\mathrm{ac(exp)}}_{\mathrm{frag}}\Big|_{g=\tilde{g}_{\mathrm{frag}}} < \mathrm{SW}^{\mathrm{cp}}_{\mathrm{frag}}\Big|_{g=0}$, then $g_E = 0$ and Fragmentation mode of entry with CP strategy will be the optimal solution as MNC does not have any incentive to deviate. In addition the pirate does not enter in this situation.

Alternatively for $S\mathrm{W}^{\mathrm{ac(exp)}}_{\mathrm{frag}}\Big|_{g=\tilde{g}_{\mathrm{frag}}} \geq \mathrm{SW}^{\mathrm{cp}}_{\mathrm{frag}}\Big|_{g=0}$, it implies $g_E = \tilde{g}_{\mathrm{frag}}$ and Fragmentation mode of entry with AC strategy will be the optimal solution.

(ii) $t > \tilde{t} \Rightarrow \prod^{cp}_{\mathrm{LDC}} > \prod^{cp}_{\mathrm{frag}}$ or $\prod^{\mathrm{ac(exp)}}_{\mathrm{LDC}}\Big|_{g=\tilde{g}_{\mathrm{LDC}}} = \prod^{cp}_{\mathrm{LDC}} \geq \prod^{cp}_{\mathrm{frag}}$

Along KT locus $\prod^{\mathrm{ac(exp)}}_{\mathrm{frag}}\Big|_{g=\tilde{g}_{\mathrm{LDC}}} = \prod^{cp}_{\mathrm{LDC}} = \prod^{\mathrm{ac(exp)}}_{\mathrm{LDC}}\Big|_{g=\tilde{g}_{\mathrm{LDC}}}$, if '*k*' and '*t*' combinations lie above the KT locus, then $\prod^{\mathrm{ac(exp)}}_{\mathrm{LDC}}\Big|_{g=\tilde{g}_{\mathrm{LDC}}} > \prod^{\mathrm{ac(exp)}}_{\mathrm{frag}}\Big|_{g=\tilde{g}_{\mathrm{LDC}}}$ which implies that FDI mode of entry with AC Strategy dominates the Fragmented mode of entry. The government chooses $g_E = \tilde{g}_{\mathrm{LDC}}$ for $\mathrm{SW}^{\mathrm{ac(exp)}}_{\mathrm{LDC}}\Big|_{g=\tilde{g}_{\mathrm{LDC}}} \geq \mathrm{SW}^{\mathrm{cp}}_{\mathrm{LDC}}\big|_{g=0}$ and $\prod^{\mathrm{fake(exp)}}_{\mathrm{LDC}} > 0$.

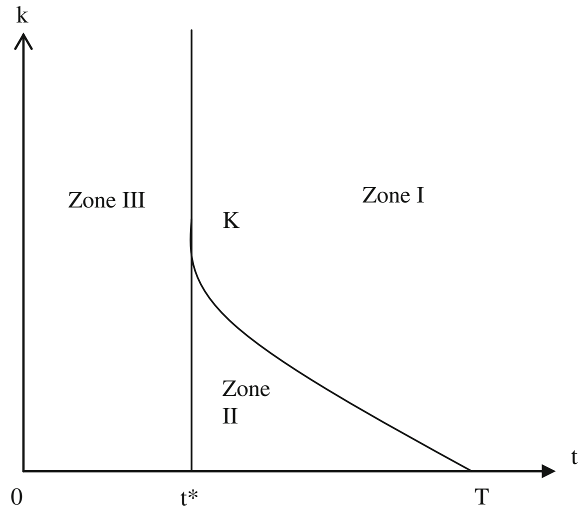
(iii) For $t > t*$ and '*k*' and '*t*' combinations lying below the KT locus, at $\tilde{g}_{\mathrm{LDC}}$ the MNC will always move to the Fragmented mode of entry as $\prod^{\mathrm{ac(exp)}}_{\mathrm{LDC}}\Big|_{g=\tilde{g}_{\mathrm{LDC}}} < \prod^{\mathrm{ac(exp)}}_{\mathrm{frag}}\Big|_{g=\tilde{g}_{\mathrm{LDC}}}$. So the only possible solution of the model will be $g_E = 0$ and FDI mode of entry with the CP strategy will be chosen. Here also the pirate will not exist.

Figure 2 illustrates the equilibrium choices for different '*k*' and '*t*' combinations.

Figure 2 defines three zones for different equilibrium configurations. In Zone I, '*t*' > t^* and '*t*' and '*k*' combinations lying above the KT locus, $\Rightarrow \prod^{\mathrm{ac(exp)}}_{\mathrm{LDC}}\Big|_{g=\tilde{g}_{\mathrm{LDC}}} = \prod^{cp}_{\mathrm{LDC}} \geq \prod^{cp}_{\mathrm{frag}}$ and $\prod^{\mathrm{ac(exp)}}_{LDC}\Big|_{g=\tilde{g}_{\mathrm{LDC}}} > \prod^{\mathrm{ac(exp)}}_{frag}\Big|_{g=\tilde{g}_{\mathrm{LDC}}}$ government chooses $g = \tilde{g}_{\mathrm{LDC}}$ for $\mathrm{SW}^{\mathrm{ac}}_{\mathrm{LDC}}\big|_{g=\tilde{g}_{\mathrm{LDC}}} \geq \mathrm{SW}^{\mathrm{cp}}_{\mathrm{LDC}}\big|_{g=0}$.

[23] $\prod^{\mathrm{ac(exp)}}_{LDC}\Big|_{g=\tilde{g}_{\mathrm{LDC}}} = \prod^{cp}_{LDC} < \prod^{cp}_{frag} = \prod^{\mathrm{ac(exp)}}_{frag}\Big|_{g=\tilde{g}_{\mathrm{frag}}}$.

Fig. 2 KT Locus and different equilibrium combinations



In Zone II, 't' > t^* but 't' and 'k' combinations lying below the KT locus $\prod_{\text{LDC}}^{\text{cp}} \geq \prod_{\text{frag}}^{\text{cp}}$ but $\prod_{\text{LDC}}^{\text{ac(exp)}}\Big|_{g=\tilde{g}_{\text{LDC}}} < \prod_{\text{frag}}^{\text{ac(exp)}}\Big|_{g=\tilde{g}_{\text{LDC}}}$, government chooses g = 0 or non-monitoring.

In Zone III for 't' > t^*, $\pi_{\text{LDC}}^{\text{cp}} \leq \pi_{\text{frag}}^{\text{cp}}$. Government chooses $g = 0$ if $\text{SW}_{\text{frag}}^{\text{ac(exp)}}\Big|_{g=\tilde{g}_{\text{frag}}} < \text{SW}_{\text{frag}}^{\text{cp}}\Big|_{g=0}$, otherwise government chooses $g = \tilde{g}_{\text{frag}}$.

5 Numerical Analysis

In this section we elaborate a numerical analysis to explain the equilibrium depicted in Proposition 5. Let us assume that

$$\begin{aligned} h(x) &= \frac{1}{x} \quad \text{for } x > 1 \\ &= 1 \quad \text{if } x = 0 \end{aligned}$$

Here $h(x)$ follows assumption A1.

Next we assume that

$$C(g) = g^2/2$$

To identify the optimal entry mode we assume different values for the parameters. The results are summarized in Table 3.

Table 3 Results of numerical analysis

Parameters	$\tilde{t}$	t and k	Profits, social welfare, optimal IPR rate	SPNE
Case 1 $A = 0$, $a = 6$, $w = 0.1$, $t = 0.4$, $F = 0.15$, $G = 0.25$ $x^{cp}_{LDC} = 3$[a] $x^{cp}_{frag} = 2.8$[b]	0.07	Let $t = 0.4$ $k = 0.9$[c]	$\prod^{cp}_{LDC} = 5.70 > \prod^{cp}_{frag} = 4.76$. (As $t > \tilde{t}$)$\tilde{g}_{LDC} = 0.5$ $\prod^{ac(exp)}_{LDC}\Big\vert_{\tilde{g}_{LDC}=0.5} = 5.7 > \prod^{ac(exp)}_{Frag}\Big\vert_{\tilde{g}_{LDC}=0.5} = 4.73$ $SW^{cp}_{LDC} = 4.225 < SW^{ac(exp)}_{LDC}\Big\vert_{\tilde{g}_{LDC}=0.5} = 6.476$ $\prod^{fake(exp)}_{LDC}\Big\vert_{\tilde{g}_{LDC}=0.5} = 0.17 > 0$	In this case as $t > \tilde{t}$ profit of FDI production mode dominates that of Fragmentation mode. Thus FDI mode of entry is chosen. Further for $\tilde{g}_{LDC} = 0.5$ Social Welfare under FDI mode is maximized for AC strategy and MNC does not have any incentive to move to Fragmentation mode of entry. Hence optimal value of monitoring rate is $g_E = \tilde{g}_{LDC} = 0.5$ with FDI production mode of entry under AC strategy becomes SPNE of the model and the fake firm enters the market
Case II $A = 0$, $a = 6$, $w = 0.1$, $t = 0.01$, $F = 0.15$, $G = 0.25$ $x^{cp}_{LDC} = 3$ $x^{cp}_{frag} = 2.8$	0.07	$t = 0.01$, $k = 0.5$	$\prod^{cp}_{LDC} = 5.70 < \prod^{cp}_{frag} = 5.87$ (As $t < \tilde{t}$)$\tilde{g}_{Frag} = .117$ $\prod^{ac(exp)}_{Frag}\Big\vert_{\tilde{g}_{fragC}=0.117} = 5.87$ $SW^{cp}_{frag} = 4.34 < SW^{ac(exp)}_{frag}\Big\vert_{\tilde{g}_{frag}=0.117} = 6.44$ $\prod^{fake(exp)}_{frag}\Big\vert_{\tilde{g}_{frag}=.117} = 0.775 > 0$	In this case $t < \tilde{t}$, thus Fragmentation profit dominates the FDI production profit. Hence Fragmentation mode of entry will be chosen. Further for $\tilde{g}_{Frag} = 0.117$ Social Welfare under fragmentation mode of entry with AC strategy is more than that for FDI Production. Hence at equilibrium the optimal value of monitoring rate is $g_E = \tilde{g}_{frag} = 0.117$ with fragmentation mode of entry and AC strategy becomes SPNE of the model and fake firm enters

[a]Under this configuration CP strategy will be a dominated strategy if $x^{cp}_{LDC} > 4.24$ otherwise $\prod^{ac(exp)}_{LDC}$ at $g = 0$ will be greater than the CP profit under FDI mode of entry. To rule out this condition we assume that $x^{cp}_{LDC} = 3$

[b]Under this configuration CP strategy will be a dominated strategy if $x^{cp}_{frag} > 4$, otherwise $\prod^{ac(exp)}_{frag}$ at $g = 0$ will be higher than the CP profit under fragmentation mode of entry. To rule out this condition we assume that $x^{cp}_{frag} = 2.8$

[c]In this case $k = 0.9$ implies that probability of copying the product of MNC by the fake firm is not very different under Fragmentation and FDI mode of entry.

Thus our paper in an effort to link the mode of entry of an MNC in an LDC market with local government choosing the IPR regime of the economy attains a situation where the LDC government may choose an IPR rate such that the MNC undergoes Fragmented mode of entry with partial or disembodied technology transfer and complete deterrence of piracy is not possible. Hence the result stands in contrast to the usual belief that the LDC government always favours FDI Production with full technology transfer.

6 Conclusion

Our model relates the mode of entry of an MNC in an LDC market to the IPR regime of the economy and tries to find out the sub-game perfect Nash equilibrium level of monitoring rate exercised by the local government under three different options of entry in the LDC market. In this process it tries to bring out the conflict of interest between a profit maximizing MNC and social welfare maximizing local LDC government.

The MNC in our model originates in an IPR protected developed country. It produces a product which can be suitably fragmented such that it can produce the product with the core technology in the DC and finish the product in the LDC. Given this possibility the MNC has three options of entry namely export, fragmented mode of entry, and FDI mode of entry. In the last two options a pirate may enter with diffuse technology. The model assumes that the MNC may adopt two possible strategies. The MNC can undertake complete copy protection strategy that completely deters the possibility of copying. Alternatively, the MNC can undertake accommodating strategy the probability of copying of the original product reduces with the level of anti-copying investment. In this case product imitation cannot be prevented completely. Further the model assumes that the government undertakes monitoring to detect the pirate if it enters the market.

Under this framework, the results show that the LDC government can induce FDI mode of entry with full technology transfer only if the transport cost of transferring the intermediate product from DC to LDC in the fragmentation mode is above a critical level. If the transport cost is above the critical level along with a relatively high probability of copying in the fragmented production, then in equilibrium, IPR monitoring cannot result in complete deterrence in piracy as the MNC adopts the AC strategy. Alternatively, if the transport cost is above the critical level but the copying probability is relatively low under fragmentation, then government chooses not to monitor and the MNC chooses FDI mode of entry with the CP strategy that results in full deterrence of piracy. However, for a low transport cost when the fragmented mode of entry is optimal, a positive monitoring rate induces the MNC to choose the AC strategy whereas if government chooses non-monitoring, the MNC adopts the CP strategy and the piracy is completely deterred.

The paper also attempts a numerical simulation analysis to actually work out possible cases where the LDC government and the MNC chooses FDI or Fragmented Production as the Sub Game Perfect Nash Equilibrium.

Thus the paper explains the situation where a MNC facing a threat of technology leakage may not undergo full technology transfer in the LDC but shift assembly line units to LDC and undertake partial technology transfer.

Appendix 1

Proof of Proposition 1

Proof Differentiating Eq. (16) w.r.t g

$$\Rightarrow \frac{\partial \bar{x}^{ac}_{frag}}{\partial g} = \frac{h'(\bar{x}^{ac}_{frag})}{h''(\bar{x}^{ac}_{frag})(1-g)} < 0 \text{ as } h'(x^{ac}_{frag}) < 0 \text{ and } h''(x^{ac}_{frag}) > 0 \text{ by assumption}$$

This implies that the MNC reduces its anti-copying investment if the local Government undertakes stronger IPR protection

Again differentiating Eq. (16) with respect to 't' the transport cost we can draw the following inference

$$\begin{aligned} &-h'(x^{ac}_{frag})(1-g)\left[\frac{(a-w-t)^2}{4} - \frac{\{a-2(w+t)\}^2}{8}\right] = 1 \\ &\Rightarrow \frac{\partial x^{ac}_{frag}}{\partial t} = \frac{h'(x^{ac}_{frag})\left(\frac{w+t}{2}\right)}{h''(x^{ac}_{frag})\left[\frac{(a-w-t)^2}{4} - \frac{\{a-2(w+t)\}^2}{8}\right]} < 0 \end{aligned} \tag{16a}$$

since the terms within the parantheses are positive and by assumption $h'(x) < 0$ and $h''(x) > 0$.

Hence Proposition 1 is proved.

Proof of the Proposition 2

For $g < g^*_{frag}$ fake firm can enjoy positive profit, and enter the market. Differentiating Eq. (18) with respect to G and t respectively

$$\frac{\delta g^*_{frag}}{\delta G} = -\frac{32\left[\{a+2(w+t)\}^2 - \frac{32F}{kh(x^{ac}_{frag})}\right]}{\left[\{a+2(w+t)\}^2 + 32G\right]^2} < 0$$

and

$$\frac{\delta g^*_{\text{frag}}}{\delta t} = \frac{128(G + F/kh(x^{\text{ac}}_{\text{frag}}))\{a + 2(w + t)\}}{\left[\{a + 2(w + t)\}^2 + 32G\right]^2} + \frac{(32Fh'(x^{\text{ac}}_{\text{frag}})/h^2(x^{\text{ac}}_{\text{frag}})\,\frac{\delta x^{\text{ac}}_{\text{frag}}}{\delta t}}{\{a + 2(w + t)\}^2 + 32G} > 0.$$

$$\frac{\partial \prod^{\text{fake}}_{\text{frag}}}{\partial g} = -\frac{\{a + 2(w + t)\}^2}{32h''(x^{\text{ac}}_{\text{frag}})}\left[h(x^{\text{ac}}_{\text{frag}})h''(x^{\text{ac}}_{\text{frag}}) - \left\{h(x^{\text{ac}}_{\text{frag}})\right\}^2\right] - G\left[h(x^{\text{ac}}_{\text{frag}}) + \frac{\left\{h(x^{\text{ac}}_{\text{frag}})\right\}^2}{h''(x^{\text{ac}}_{\text{frag}})(1 - g)}\right] < 0$$

where $h'(x^{\text{ac}}_{\text{frag}}) < 0$ and $h''(x^{\text{ac}}_{\text{frag}}) > 0$ by assumption.

The above expression is unambiguously negative if A1 holds.

References

Banerjee DS (2003) Software piracy: a strategic analysis and policy instruments. Int J Ind Organ 21:97–127

Banerjee DS (2011) On the sufficiency of regulatory environment in combating piracy. J Regul Econ 40:160–176

Banerjee D, Banerjee T, Raychouhuri A (2008) Optimal enforcement and anti-copying strategies to counter copyright infringement. Jpn Econ Rev 59(4):519–535

Eicher T, Kang JW (2005) Trade FDI or acquisition: optimal entry modes for multinational. J Dev Econ 77:207–208

Kiema I (2008) Commercial piracy and intellectual property policy. J Econ Behav Organ 68:304–318

Lin P, Saggi K (1999) Incentives for foreign direct investment under imitation. Canadian J Econ 32:1275–1298

Long NV (2005) Outsourcing and technology spillovers. Int Rev Econ Finan 14:297–304

Lu Y, Poddar S (2011) Accommodation or deterrence in the face of commercial Piracy: the impact of IPR protection. Oxford Economic Papers

Lu Y, Poddar S (2012) Accommodation or deterrence in the face of commercial piracy: the impact of IPR protection. Oxford Econ Pap 64(3):518–538

Mattoo A, Olarreaga M, Saggi K (2004) Mode of foreign entry, technology transfer and FDI policy. Policy Research Working Paper 2737

Pack Howard, Saggi K (2001) Vertical technology transfer via international outsourcing. J Dev Econ 65:398–415

Poddar S (2005) Why software piracy rates differ—a Theoretical Analysis Working Paper No. 0515, Dept. of Economics, National University of Singapore

Silve J, Bernhardt D (1998) Pirated for profit. Canadian J Econ 31:886–899

Vishwasrao S (1994) Intellectual property rights and the mode of technology transfer. J Dev Econ 44:381–402

Zigic K (2000) Strategic trade policy, intellectual property rights protection and North-South trade. J Dev Econ 61:27–60

Part II
International Trade and Institutions

International Trade and the Size of the Government

Rajat Acharyya

Abstract This paper examines the relationship between trade openness and size of the government. Under homothetic taste, and thus without the real income effect, a more open (small) economy may indeed have a larger size of the government even when the public good that it produces is not traded. When the public good is the most labour intensive good, the absolute size of the government expands with a tariff reduction under a reasonable assumption regarding employment shares of the export good. An additional condition requiring a sufficiently small, although not necessarily less than unity, value of the price elasticity of demand for the public good ensures that the relative size of the government expands as well.

Keywords Government size · Openness · Tariff reduction · Non-traded good · Externality

JEL Classification F11 · H11 · H42

1 Introduction

Growing concerns over huge budgetary and fiscal deficits in a large number of countries have inspired quite a few researchers in recent times to investigate the determinants of the size of the government. External risks and political instability are the two major explanations that are put forward for larger size of the government measured by the share of government expenditure in gross domestic product. Rodrik (1998) argues quite convincingly that government spending provides social insurance in open economies that are subject to external shocks and risks due to

R. Acharyya (✉)
Department of Economics, Jadavpur University, Calcutta 700032, India
e-mail: rajat.acharyya@gmail.com

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_5

terms of trade volatility. This is the reason we observe larger governments in more open economies.[1] Annett (2001), on the other hand, argues that higher ethnolinguistic and religious fractionalization within an economy increases the political risk for an incumbent government. This induces the government to pacify the excluded interest groups by increasing the level of government consumption.[2] Thus, whereas Rodrik finds government expenditure as means of lowering external risks for an open economy, Annett finds it as means of lowering political risk and instability.

Measuring the size of the government by the share of expenditure on public good in the total expenditure on the private and public goods, Anwar and Zheng (2004) show that the size of the government is determined by the availability of the primary factors of production. In their two-sector, two-factor general equilibrium model of a closed economy, the public good is essentially a private good produced by the government and sold at the marginal cost. In such a framework, they show that an increase in the supply of capital (for example, through an inflow of foreign capital) lowers the size of the government sector (or the share of expenditure on the public good) if the production of the public good is relatively labour intensive.

An interesting implication of this result that the size of the government sector depends on factor endowments is similar to the Heckscher-Ohlin-Samuelson (hereafter HOS) pattern of *relative* size of different sectors and, therefore, of the comparative advantage of the economies endowed with different amounts of the primary factors. Their analysis thus creates scope for free international trade to have a similar impact on the size of the government sector as the international factor mobility when both the private and the public goods are traded in a standard HOS structure. This is a simple logical derivation from the celebrated Factor Price Equalization theorem. International trade by freeing up resources from the import-competing sectors allows production of exports to increase. If the public good is exportable, then it is straightforward to argue that the size of the government sector, both the absolute size and the share of expenditure, should increase with the opening up of trade. Thus the observed larger government size in open economies may be due to changes in effective factor supplies that free international trade brings in, provided, of course, the public good is exportable.

But such a result does not follow immediately in a standard HOS framework of an economy if the public good is *not* a traded good. It is thus non-trivial to examine the relationship between openness to external trade and the size of the government. This is the primary concern of this paper. More precisely, I examine how freer international trade affects the size of the government through changes in the effective supplies of the primary factors of production, when the private good that the government produces is a non-traded good. In an extended HOS framework of a small open economy, I consider two privately produced traded goods and one government produced non-traded good (henceforth loosely called public good to

---

[1]Alesina and Wacziarg (1998) also have found that the government size is larger in smaller economies and the smaller economies are the ones usually more open to the external trade.

[2]Similar logic can be found in Blomberg (1996) and Velasco (1997).

differentiate it with the two goods produced in the private sectors), all using the same resources—labour and capital. A tariff reduction in such a context is shown to expand the absolute size of the government (defined as the expenditure on the public good), under a set of fairly general conditions. For example, if the export good is labour intensive relative to the import-competing good, by the standard price-magnification effect, a tariff reduction raises the money wage relative to the rate of return to capital. This change in relative wage raises the price of the non-traded public good if it, in turn, is labour intensive relative to the export good. Production of the public good, on the other hand, falls by the consequent decline in its *relative* demand, but increases because tariff reduction raises, for certain values of the employment shares in the export sector, the output of the import-competing good which necessitates a proportionate increase in the production of the public good under homothetic tastes to clear the domestic market. This market-clearing expansionary effect appears to be stronger than the contractionary demand effect so that the production of the public good increases at the end. Accordingly, with both the price and the output of the non-traded good increasing following a tariff reduction, the absolute size of the government increases. The relative size of the government, defined as the share of total expenditure spent on the public good, on the other hand, increases when the ratio of the share of export production to that of the import-competing production in national income exceeds the ratio employment shares of the private traded sectors. Otherwise, the expansion of the relative size of the government depends on the value of the price elasticity of demand for the public good. However, if the public good is price-inelastic, the government size increases regardless of whether the share of export production is large or small. Since lower levels of tariffs essentially make an economy more open to the external world, these results should mean that, ceteris paribus, more open economies have larger governments as is in fact observed by Alesina and Wacziarg (1998) and Rodrik (1998).

The production of the non-traded public good may generate positive externality effects in the rest of the economy. If the public good is in the nature of physical and/or social infrastructure (such as hospitals or health care, building of paved roads and the like), such externality effects may be in the nature of improvement in productivity of labour and capital. When such productivity effects of the production of the public good are taken into account, a tariff reduction still may increase the size of the government sector but now under more stringent conditions.

The rest of the paper is organized as follows. In Sect. 2 we discuss the framework of our small open economy. Section 3 works out the effects of tariff reduction on the production of the public good and the size of the government and Sect. 4 considers the positive externality effect of the production of the public good. Finally, concluding remarks are provided in Sect. 5.

2 The Framework of a Small Open Economy

Consider a small open economy with two traded goods, X and Y, produced in the private competitive sectors and a non-traded public good G produced by the government. Though we label G as a public good, following Anwar and Zheng (2004), we assume that it has the same characteristics as the private goods.[3] The only difference lies in the pricing of these goods. Whereas the marginal cost pricing for the two private goods are outcomes of perfect competition, the marginal cost pricing for the public good is the outcome of no-profit-no-loss motive of the government. Given the nature of the good produced by the government, we abstract from any positive externality that such production may generate. However, if instead the good under consideration had been in the nature of a semi or pure public good, we would have expected it to generate a positive externality in production. In a later section, we will explore whether and how such a dimension of the good produced by the government determines the size of the government sector.

All the three goods are produced by labour (L) and domestic capital (K) under constant returns-to-scale technology. The world prices of the traded goods are given to our small open economy and there is an initial ad-valorem tariff t on the import good Y. Suppose, the export good X is not domestically consumed, and that the consumers allocating their budget on goods Y and G have homothetic taste. Thus, the optimal consumption ratio, $\frac{G^d}{Y^d}$, depends only on the relative price of the public good. This assumption helps us ignore the income effects of a tariff reduction. Note that the tariff reduction that we consider here will raise the real income of our small open economy by reducing the distortion that a tariff introduces. As long as the public good is normal in consumption, such a real income increase will raise the demand for it and, therefore, its production and the size of the government. Thus by assuming homothetic taste we rule out this favourable effect to examine whether greater openness can still lead to a larger size of the government.

Competition in the private sector means prices equal average (and marginal) cost of production for the traded goods

$$P_X^* = a_{LX}W + a_{KX}r \tag{1}$$

$$(1+t)P_Y^* = a_{LY}W + a_{KY}r \tag{2}$$

where prices with asterisk denote the world prices.

[3]In mixed economies like India, the government is not only the provider of pure and semi public goods, but also of many private goods which cannot even be justified on grounds of positive externalities as, for example, in cases like public transport, public education, or health care. It is the size of this component of the government activities that really is a concern for many. This is one reason why we confine ourselves with only a publicly provided private good instead of a pure public good.

Government, on the other hand, though is the sole producer of the public good, charges the price which covers the average cost under the no-profit-no-loss motive:

$$P_G = a_{LG}W + a_{KG}r \tag{3}$$

There are also the least-cost choices of input coefficients a_{ij}, which under the constant returns-to-scale technology depends only on the factor prices

$$a_{ij} = a_{ij}(W/r),\ i = L,\ K;\ j = X,\ Y,\ G \tag{4}$$

Since the public good is a non-traded good, its market must clear domestically. Given homothetic taste, we must have then the following market-clearing condition

$$\frac{G^d}{Y^d} = f\left(\frac{P_G}{P_Y}\right) = \frac{G}{Y} \tag{5}$$

where, G and Y are the levels of output of the public good and the import-competing good respectively. We close the model with the following full employment conditions

$$\overline{L} = a_{LX}X + a_{LY}Y + a_{LG}G \tag{6}$$

$$\overline{K} = a_{KX}X + a_{KY}Y + a_{KG}G \tag{7}$$

The determination of equilibrium configuration of factor prices, price and output of the non-traded public good, and the output of the traded private goods is fairly standard as in a typical HOS model. By the competitive conditions (1) and (2) and the least-cost choices (4), the money wage and the rate of return to capital are uniquely determined by the given world prices of the two private traded goods, for any given rate of the ad-valorem tariff. By the marginal cost pricing rule for the public good as described in (3), the average cost then determines the price of the public good. Finally, given the least-cost input choices, the market-clearing condition (5) together with the full employment conditions (6) and (7) determine the three output levels.

In the rest of our analysis, we assume that the export good G is the most labour intensive and the import-competing good is the least labour intensive. That is,

$$\frac{a_{LG}}{a_{KG}} > \frac{a_{LX}}{a_{KX}} > \frac{a_{LY}}{a_{KY}} \quad \forall \frac{W}{r} \tag{8}$$

The implication of this assumption will be made clear later. This intensity ranking is not too unrealistic in the context of the nature of exports and imports of the developing countries in general. On the other hand, most of the production in the government sector is labour intensive with a view of generating employment opportunities.

3 Openness and the Size of the Government

In this framework, the level of tariff indicates the extent to which our small economy is open to the external trade. The lower the tariff on imports, more open the economy is. Thus if a ceteris paribus tariff reduction can be shown to raise the share of the public good in total expenditure of the economy, we can conclude that more open economies have larger governments. This is the thought experiment we carry out in this section.

Since we have assumed in (8) that imports are capital intensive relative to exports, by the price-magnification effect, a tariff reduction raises the money wage and lowers the rate of return to capital

$$\widehat{W} = -\frac{\theta_{KX}}{|\theta|_{XY}}\widehat{T},\ \hat{r} = \frac{\theta_{LX}}{|\theta|_{XY}}\widehat{T} \tag{9}$$

where hat over a variable denote the proportional change; θ_{ij} is the cost share of input-i in sector-j; $|\theta|_{XY} = \theta_{LX}\theta_{KY} - \theta_{LY}\theta_{KX} > 0$; and $\widehat{T} \equiv \frac{d(1+t)}{(1+t)}$.

The price of the public good, on the other hand, rises with the tariff reduction by our assumption that the public good is the most labour intensive or least capital intensive (so that $\theta_{KX} > \theta_{KG}$)

$$\widehat{P}_G = -\frac{\theta_{KX} - \theta_{KG}}{|\theta|_{XY}}\widehat{T} \tag{10}$$

With the domestic price of imports falling, this means a switch in the domestic demand in favour of the import good Y and away from the public good. Since output of the public good is demand-determined by virtue of it being a non-traded good, its production must therefore contract on account of this demand effect of a tariff reduction. There will be, in addition, a market-clearing effect on the production of the public good necessitated by the homothetic taste assumption. As evident from (5),

$$\widehat{G} = \frac{(\theta_{KY} - \theta_{KG})\varepsilon_G}{|\theta|_{XY}}\widehat{T} + \widehat{Y} \tag{11}$$

where, $\varepsilon_G = -\frac{\widehat{G}-\widehat{Y}}{\widehat{P}_G-\widehat{P}_Y}$ is the (absolute) value of the price elasticity of (relative) demand for the non-traded good.

The first term captures the demand effect induced by the change in the relative price of the public good following the tariff reduction. The second term captures the market-clearing effect under the assumption of homothetic taste. Under such an assumption, the relative demand for the public good can change only when the relative price of it changes. Thus, if a tariff reduction raises the output of the import-competing good, then for any given relative price, the production of the public good must also increase proportionately to match the (relative) demand for the public good and, therefore, to clear the domestic market for the non-traded good.[4] Note by the assumption that our economy is small, any change in production of the traded goods cannot affect the world prices and hence the price of the public good relative to the (domestic) price of imports.

Thus, to determine how the production of the public good changes with a tariff reduction, we need to know how the outputs of the traded private goods change. First of all, these output levels would change by the factor substitution effect. The increase in the wage level relative the rate of return to capital makes all lines of production relatively more capital intensive. With labour and capital fully employed, the output levels must, therefore, adjust. Had all the three goods been traded, the output levels would have changed only by these factor substitution effects. But with the public good being non-traded, there will be further changes in the output of the traded private goods. As spelled out above, for any given output level of the import-competing good, the production of the public good falls with the tariff reduction. The consequent release of some labour and capital previously employed in the government sector, therefore, increases the *net* endowment of these factors for the private sub-economy. This triggers the output magnification effects. However, given the three goods, the direction of change in the output levels triggered by the factor substitution and net endowment changes will not follow directly from the particular intensity ranking assumed in (8). Formally, as shown in the appendix

$$\widehat{Y} = -\left[\frac{\eta\lambda_{KX} - |\lambda|_{XG}(\theta_{KY} - \theta_{KG})\varepsilon_G}{(\lambda_{KX} - \lambda_{LX})|\theta|_{XY}}\right]\widehat{T} \tag{12}$$

$$\widehat{X} = \left[\frac{\lambda_{KX}\delta(\lambda_{KX} - \lambda_{LX}) + (1 - \lambda_{KX})\eta}{(\lambda_{KX} - \lambda_{LX})\lambda_{KX}|\theta|_{XY}} + \frac{\{(1 - \lambda_{KX})|\lambda|_{XG} + (\lambda_{KX} - \lambda_{LX})\lambda_{KG}\lambda_{KX}\}(\theta_{KG} - \theta_{KY})\varepsilon_G}{(\lambda_{KX} - \lambda_{LX})\lambda_{KX}^2|\theta|_{XY}}\right]\widehat{T} \tag{13}$$

$$\widehat{G} = -\left[\frac{\eta\lambda_{KX} + (\theta_{KY} - \theta_{KG})|\lambda|_{XG}\varepsilon_G}{(\lambda_{KX} - \lambda_{LX})|\theta|_{XY}}\right]\widehat{T} \tag{14}$$

where, $\eta \equiv \lambda_{LX}\sigma_X + \left(\lambda_{LX}\frac{\lambda_{KY}}{\lambda_{KX}}\theta_{LY} + \lambda_{LY}\theta_{KY}\right)\sigma_Y + \left(\lambda_{LX}\frac{\lambda_{KG}}{\lambda_{KX}}\theta_{LG} + \lambda_{LG}\theta_{KG}\right)\sigma_G > 0;$

$\delta \equiv \frac{1}{\lambda_{KX}}[\lambda_{KX}\theta_{LX}\sigma_X + \lambda_{KY}\theta_{LY}\sigma_Y + \lambda_{KG}\theta_{LG}\sigma_G] > 0;$

$\sigma_j = \frac{\hat{a}_{Kj} - \hat{a}_{Lj}}{(\widehat{W} - \hat{r})},\ j = X,\ Y,\ G,$ is the factor substitution elasticity in sector-j;

$|\lambda|_{XG} = \lambda_{LX}\lambda_{KG} - \lambda_{LG}\lambda_{KX} < 0$

[4]In case of non-homothetic tastes, this market-clearing effect would not have been there. Instead, the output of the public good would have been determined, in addition to the demand effect induced by change in price, by the demand effect consequent upon change in the real income.

The first term in the parenthesis of each of the expressions in (12) and (13) captures the factor substitution effect whereas the second term captures the demand-induced net endowment-change effect as spelled out above. First of all, given the intensity ranking in (8), the output of the import-competing good rises, $\widehat{Y} > 0$, if,

$$\lambda_{KX} > \lambda_{LX} \tag{15}$$

Given (15), both the factor substitution effect and the net endowment-change effect are favourable for the import-competing sector.[5] The factor substitution effect, on the other hand, lowers the output of the export good X whereas the other effect is ambiguous.

In rest of our analysis, we shall assume (15) in addition to (8) to reduce the multiplicity of cases. Thus, given these assumptions, a tariff reduction raises the production of the import-competing good whereas that of the export good may fall.[6]

Finally, since the output of the import-competing good increases, $\widehat{Y} > 0$, from (11) it appears that the market-clearing effect raises the production of the public good and thus works in the opposite direction of the demand effect.

Therefore,

Proposition 1 *By (8) and (15), a tariff reduction* $(\widehat{T} < 0)$ *raises the volume of production of good G and the absolute size of the government.*

Proof By (8) and (15), $|\theta|_{XY} > 0$ and $\theta_{KG} < \theta_{KY}$. Then, it is immediate from (14) that $\widehat{G} > 0$. This completes the first part of the proof.

For the second part, define the expenditure on good G as the absolute size of the government

$$S = P_G G \tag{16}$$

Thus, by (8), $\widehat{S} = \widehat{P}_G + \widehat{G} > 0$.

Hence the claim. □

A tariff reduction affects the absolute size of the government in three ways. First is the price effect. Tariff reduction raises the price of the public good and thus the expenditure on it for any given consumption and output levels. The other two effects

---

[5]This result that a fall in tariff actually raises the domestic production of the import-competing good is similar to the well-known Metzler Paradox in the literature on international trade. This, however, arises due to the existence of a non-traded good which uses the same factors of production as do the traded goods.

[6]The imports, M, can still increase. Since $M = Y_d - Y$, we can write $\widehat{Y}_d = \rho\widehat{Y} + (1-\rho)\widehat{M}$, where $\rho = \frac{Y}{Y_d}$. Since Y_d increases, even if the output Y increases faster, by the property of weighted average, we have $\widehat{Y} > \widehat{Y}_d > \widehat{M}$. Thus $\widehat{M}$ may still be positive.

are the adverse demand effect and the favourable market-clearing effect on the output of the public good as spelled out earlier. By assumptions in (8) and (15), it turns out that the market-clearing effect outweighs the adverse demand effect. The output of the public good, therefore, rises at the end. Consequently, the initial expansion of the absolute size of the government through the price effect gets magnified.

How does a tariff reduction change the relative size of the government? Following the literature, the relative size is defined as the expenditure on public good as a proportion of the national income (or total expenditure)

$$S_R = \frac{P_G G}{P_G G + P_X^* X + P_Y^* Y} \tag{17}$$

The (proportional) change in the relative size of the government thus can be written as,

$$\widehat{S}_R = (1 - \theta_G)[\widehat{P}_G + \widehat{G}] - \theta_X \widehat{X} - \theta_Y \widehat{Y} \tag{18}$$

where θ_j, $j = G, X, Y$, is the share of sector-j in national income and $\sum_j \theta_j = 1$. Substitution of values from (10), (12)–(14) yields,

$$\widehat{S}_R = \frac{\begin{array}{c}-[\{(\theta_X + \theta_Y)(\theta_{KX} - \theta_{KG})\lambda_{KX} + \theta_X \lambda_{KX} \delta\}(\lambda_{KX} - \lambda_{LX}) + \theta_X \eta] \\ + (\theta_{KY} - \theta_{KG})[\theta_Y(\lambda_{KX} - \lambda_{LX}) - \theta_X(\lambda_{KY} - \lambda_{LY}]\varepsilon_G\end{array}}{(\lambda_{KX} - \lambda_{LX})\lambda_{KX}|\theta|_{XY}} \widehat{T} \tag{19}$$

Following Lemma is useful in determining the change in the relative size of the government

Lemma 1 *By* (8) *and* (*15*), $\lambda_{KY} > \lambda_{LY}$.

Proof: By (8),

$$\frac{\lambda_{LX}}{\lambda_{KX}} > \frac{\lambda_{LY}}{\lambda_{KY}}$$

which by (15) implies,

$$\frac{\lambda_{KY}}{\lambda_{LY}} > \frac{\lambda_{KX}}{\lambda_{LX}} > 1 \tag{20}$$

Hence the claim. □

It is immediate from (19) and Lemma 1 that since a tariff reduction means $\widehat{T} < 0$, so $\widehat{S}_R > 0$ if,

$$\frac{\theta_X}{\theta_Y} > \frac{\lambda_{KX} - \lambda_{LX}}{\lambda_{KY} - \lambda_{LY}} \tag{21}$$

But if $\frac{\theta_X}{\theta_Y}$ is smaller than the critical value defined in (21), then the coefficient of ε_G in (19) is positive, in which case $\widehat{S}_R > 0$ if,

$$\varepsilon_G < \frac{\{(\theta_X + \theta_Y)(\theta_{KX} - \theta_{KG}) + \theta_X\}(\lambda_{KX} - \lambda_{LX})\lambda_{KX} + \theta_X\eta}{(\theta_{KY} - \theta_{KG})[\theta_Y(\lambda_{KX} - \lambda_{LX}) - \theta_X(\lambda_{KY} - \lambda_{LY})]} \tag{22}$$

Therefore,

Proposition 2 *Given (8) and (15), a more open economy will have a larger size of the government if the ratio of share of the export sector to that of the import-competing sector in national income is sufficiently large in the sense defined in (21). Otherwise, a smaller value of the price elasticity of demand ensures that a more open economy will have a larger government.*

Proof: Follows from the above discussion. □

The reason for these conditions is simple. As we have spelled out, given our factor intensity assumptions a more open economy in the sense of imposing a lower tariff on imports, will have a larger absolute size of import-competing sector whereas may have a smaller absolute size of the export sector. The relative size of the government as defined in (17) above therefore increases if the import-competing sector has a smaller share in the national income of our small open economy in the sense defined in (21). But if the share of the import-competing sector is larger, the size of the private sub-economy is expanding. Thus for the relative size of the government to increase, we need that this increase in the (absolute) size of the private sub-economy is small enough. Since, it is the fall in the demand for the public good and consequent increase in the net availability of labour and capital that cause expansion of the production of both the private traded goods, we must have this demand effect smaller. This is in fact captured by the condition in (22).

Figure 1 below illustrates the results summarized in Propositions 1 and 2, where ε_G^* is the critical value of the price elasticity defined in the right-hand-side of (22).

Finally, it is interesting to note that the critical value of the price elasticity defined in (22) is greater than unity if (see Appendix),

$$\theta_{KX} > \frac{\theta_X}{\theta_X + \theta_Y}\theta_{KG} + \frac{\theta_Y}{\theta_X + \theta_Y}\theta_{KY} \equiv \theta_{KX}^* \tag{23}$$

Fig. 1 Size of the government when $\frac{\theta_X}{\theta_Y}$ is small

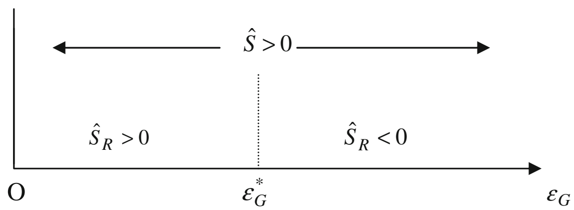
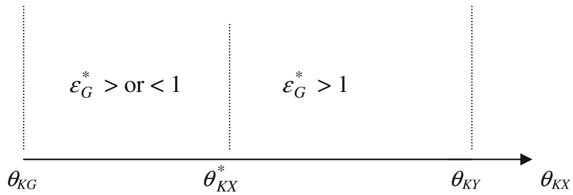


Fig. 2 Capital cost share and the critical value of price elasticity



By (8) and the property of a weighted average, with weights adding up to unity, $\theta_{KG} < \theta_{KX}^* < \theta_{KY}$. Thus θ_{KX} can be larger than θ_{KX}^* and still satisfy the assumed intensity ranking in (8). Of course, since (23) is just a sufficient condition, the critical value of ε_G may exceed unity even when is θ_{KX} smaller. This is illustrated in Fig. 2.

Hence, the size of the government can be larger in a more open economy even if the non-traded public good that it produces is price elastic in demand.

All these results together imply the following

Corollary 1 *Given (8), (15) and (23), a more open economy has a larger size of the Government regardless of the value of* $\frac{\theta_X}{\theta_Y}$*, if the demand for the public good it produces is price inelastic.*

This can be explained as follows. Note that if $\frac{\theta_X}{\theta_Y}$ is large in the sense defined in (21), the relative size of the government expands regardless of how large is the price elasticity of demand for the non-traded public good. On the other hand, if $\frac{\theta_X}{\theta_Y}$ is smaller, under (23), the relative size of the government expands if the public good that it produces is not highly price elastic in demand. These two results together imply that if the capital cost share in the export sector satisfies (23), and the public good is price-inelastic in demand, the relative size of the government increases regardless of whether $\frac{\theta_X}{\theta_Y}$ is large or small.

4 Robustness: Positive Externality

The goods and services produced by the government often generate positive externalities for other sectors or agents. Better roads or physical infrastructure, for example, improves factor productivity and reduce production costs for goods produced in the rest of the economy. Similarly, health care services are likely to improve productivity of workers. One might wonder then how the results derived above would have changed if such positive *production* externalities had been taken into account. To examine this in the simplest possible manner, suppose marginal productivities of labour and capital improves with larger production (and consumption) of good G, though not uniformly. That is, *for any given wage-rental ratio*, less number of workers and smaller amount of capital will be required to produce one unit of each good when the production of good G increases. The least-cost choice of input as specified will still be relevant but now that will reflect factor substitution *for any given production (and consumption) of good G*. However, we assume that the productivity of workers and capital employed in sector G does not change. This may not necessarily be the case, but this assumption makes the algebra less cumbersome. To capture these *external effects*, we rewrite the input requirement condition (4) as

$$a_{ij} = a_{ij}(W/r, G),\ i = L, K;\ j = X, Y \tag{4a}$$

$$a_{iG} = a_{iG}(W/r),\ i = L, K \tag{4b}$$

The change in per unit requirement of labour in sector X, for example, then comes from two sources: change in the wage-rental ratio, and change in the volume of production of good G:

$$\hat{a}_{LX} = -\eta_{LX}(\widehat{W} - \hat{r}) - e_{LX}\widehat{G} \tag{24}$$

where, $\eta_{LX} \equiv -\frac{(w/r)}{a_{LX}}\frac{\partial a_{LX}}{\partial(W/r)}$ is the absolute factor price elasticity of labour demand per unit of output, and $e_{LX} \equiv -\frac{G}{a_{LX}}\frac{\partial a_{LX}}{\partial G} > 0$ is the percentage decline in per unit labour requirement following one percent increase in the production of good G, which captures the positive externality in terms of gain in factor productivities. The changes in other factor coefficients can similarly be specified. To simplify matters further, we assume that productivity increase is uniform for all workers in whichever private sector they are employed such that $e_{LX} = e_{LY} = e_L$. Similarly, we assume $e_{KX} = e_{KY} = e_K$.

The positive externality captured this way produces two additional effects that are important for examining its impact on the size of the government. First is the scale effect. Since per unit labour and capital requirement now declines, at initial output levels of good X and Y, more labour and capital will now be available. Such increases in the effective supply of labour and capital lead to changes in the production of the three goods. At the initial wage-rental ratio, we have scale expansion or contraction. However, the direction of changes in the output levels depend on the magnitude of increases in labour productivity vis-à-vis that in capital productivity. If labour productivity increases less than the capital productivity, for example, the effective relative supply of labour *declines*. For any given output level of good G, in the private sector of the economy we can then expect the production of relatively labour intensive good X to fall, similar to the output magnification effect at initial wage-rental ratio. Such fall in production of good X makes labour available in excess of what may be required to expand the production of good Y. This excess supply of labour in turn creates scope for an expansion of the production of good G.

The second additional effect is the technique effect induced by the positive externality generated by the production of good G. Productivity increase raises the money wage and the rate of return to capital proportionately. The wage-rental ratio thus increases if labour productivity rises more than proportionately than the increase in capital productivity. Algebraically, as shown in the appendix,

$$\widehat{W} - \hat{r} = (e_L - e_K)\widehat{G} - \frac{1}{|\theta|_{XY}}\widehat{T} \tag{25}$$

The change in the wage-rental ration induces a technique effect and brings in further changes in the output levels. An increase in the wage-rental ratio, for example, substitutes labour by capital in all lines of production. The consequent availability of labour per unit of output expands the output of the relatively labour intensive good *X* vis-à-vis good *Y*, again similar to the output magnification effect. The net availability of labour for production of good *G* thus falls and accordingly this good being the most labour intensive by assumption, the production of it declines. But if capital productivity increased more than the labour productivity, the production of good *G* would have increased.

What appears from the above discussion is that the positive externality generated by the production of good *G* as captured by the productivity effects causes an increase in the production of good *G* and an unambiguous expansion of the size of the government sector if labour productivity increase is not larger than the capital productivity increase. Note that the price of good *G* rises unambiguously since both the labour cost and the capital cost increase due to the externality effect. Algebraically all these can be verified from the following expression (see appendix)

$$\widehat{P}_G = (\theta_{LG}e_L + \theta_{KG}e_K)\widehat{G} - \frac{\theta_{KX} - \theta_{KG}}{|\theta|_{XY}}\widehat{T} \tag{26}$$

$$\widehat{G} = -\left[\frac{\eta\phi\lambda_{KX} + (\theta_{KY} - \theta_{KG})|\lambda|_{XY}\varepsilon_G}{(\lambda_{KX} - \lambda_{LX})|\theta|_{XY} - \left\{\eta(e_L - e_K)\lambda_{KX} + \psi\lambda_{KX} + (\phi - 1)|\lambda|_{XY}\right\}|\theta|_{XY}}\right]\widehat{T} \tag{27}$$

where, $\psi = (\lambda_{LX} + \lambda_{LY})e_L - \frac{(\lambda_{KX} + \lambda_{KY})\lambda_{LX}}{\lambda_{KX}}e_K$ and $\phi = 1 + (\theta_{LG}e_L + \theta_{KG}e_K) > 1$.

As shown in the appendix, $\psi < 0$ for $e_L \leq e_K$. Accordingly, given the assumption in (15), $|\lambda|_{XY} > 0$ and $|\theta|_{XY} > 0$, an increase in the production of good *G* after tariff reduction is more likely when $e_L \leq e_K$. Since the productivity increase raises the price of good *G* unambiguously, so in such a case the absolute size of the government sector (as measured by $P_G G$) rises as well.

The changes in the conditions for an increase in the relative size of the government can similarly be worked out.

In sum, increased trade causing a larger size of the government sector is plausible even when the good produced by the government generates positive production externality by raising factor productivities. However, size expansion will be less if there had been no positive externality.

5 Conclusion

What we have established in this paper is that under homothetic taste, and thus even without the real income effect, a more open (small) economy may indeed have a larger size of the government even when it produces a non-traded public good. The absolute size of the government expands under a reasonable set of assumptions regarding factor intensity and employment shares of the goods. For a larger relative size of the government, we require an additional condition of either a higher share of the export sector in national income or a sufficiently small, although not necessary less than unity, value of the price elasticity of demand for the public good.

Appendix

1. Changes in the output levels.

From the full employment conditions (6) and (7) in the text, we can write:

$$\lambda_{LX}(\widehat{X}+\hat{a}_{LX})+\lambda_{LY}(\widehat{Y}+\hat{a}_{LY})+\lambda_{LG}(\widehat{G}+\hat{a}_{LG})=0 \tag{28}$$

$$\lambda_{KX}\left(\widehat{X}+\hat{a}_{KX}\right)+\lambda_{KY}\left(\widehat{Y}+\hat{a}_{KY}\right)+\lambda_{KG}\left(\widehat{G}+\hat{a}_{KG}\right)=0 \tag{29}$$

Solving for $\widehat{X}$ from (29) yields,

$$\widehat{X}=-\hat{a}_{KX}-\frac{\lambda_{KY}}{\lambda_{KX}}\left(\widehat{Y}+\hat{a}_{KY}\right)-\frac{\lambda_{KG}}{\lambda_{KX}}\left(\widehat{G}+\hat{a}_{KG}\right) \tag{30}$$

Substituting (30) in (28) we get:

$$\begin{aligned}
&\lambda_{LX}\hat{a}_{LX}+\lambda_{LX}\left[-\hat{a}_{KX}-\frac{\lambda_{KY}}{\lambda_{KX}}\left(\widehat{Y}+\hat{a}_{KY}\right)-\frac{\lambda_{KG}}{\lambda_{KX}}\left(\widehat{G}+\hat{a}_{KG}\right)\right]\\
&\quad+\lambda_{LY}\left(\widehat{Y}+\hat{a}_{LY}\right)+\lambda_{LG}\left(\widehat{G}+\hat{a}_{LG}\right)=0\\
\Rightarrow\quad&\lambda_{LX}(\hat{a}_{LX}-\hat{a}_{KX})+\left[\lambda_{LY}-\lambda_{LX}\frac{\lambda_{KY}}{\lambda_{KX}}\right]\widehat{Y}-\lambda_{LX}\frac{\lambda_{KY}}{\lambda_{KX}}\hat{a}_{KY}+\left[\lambda_{LG}-\lambda_{LX}\frac{\lambda_{KG}}{\lambda_{KX}}\right]\widehat{G}\\
&\quad-\lambda_{LX}\frac{\lambda_{KG}}{\lambda_{KX}}\hat{a}_{KG}+\lambda_{LY}\hat{a}_{LY}+\lambda_{LG}\hat{a}_{LG}=0\\
\Rightarrow\quad&\lambda_{LX}(\hat{a}_{LX}-\hat{a}_{KX})-\frac{|\lambda|_{XY}}{\lambda_{KX}}\widehat{Y}-\lambda_{LX}\frac{\lambda_{KY}}{\lambda_{KX}}\hat{a}_{KY}-\frac{|\lambda|_{XG}}{\lambda_{KX}}\widehat{G}\\
&\quad-\lambda_{LX}\frac{\lambda_{KG}}{\lambda_{KX}}\hat{a}_{KG}+\lambda_{LY}\hat{a}_{LY}+\lambda_{LG}\hat{a}_{LG}=0
\end{aligned} \tag{31}$$

Using $\widehat{G} = \frac{(\theta_{KY}-\theta_{KG})\varepsilon_G}{|\theta|_{XY}}\widehat{T} + \widehat{Y}$ from Eq. (11) in the text this boils down to,

$$\begin{aligned}\lambda_{LX}(\hat{a}_{LX} - \hat{a}_{KX}) - \left[\frac{|\lambda|_{XY} + |\lambda|_{XG}}{\lambda_{KX}}\right]\widehat{Y} - \lambda_{LX}\frac{\lambda_{KY}}{\lambda_{KX}}\hat{a}_{KY} - \frac{\varepsilon_G(\theta_{KY} - \theta_{KG})|\lambda|_{XG}}{|\theta|_{XY}\lambda_{KX}}\widehat{T} \\ - \lambda_{LX}\frac{\lambda_{KG}}{\lambda_{KX}}\hat{a}_{KG} + \lambda_{LY}\hat{a}_{LY} + \lambda_{LG}\hat{a}_{LG} = 0\end{aligned} \tag{32}$$

Using the least-cost conditions, $\theta_{Lj}\hat{a}_{Lj} + \theta_{Kj}\hat{a}_{Kj} = 0$, $\sum_j \lambda_{ij} = 1$, and recalling from the text that $\sigma_j = \frac{\hat{a}_{Kj}-\hat{a}_{Lj}}{(\widehat{W}-\hat{r})}$, $j = X, Y, G$, is the factor substitution elasticity in sector-j, (32) boils down to,

$$\begin{aligned}&\lambda_{LX}(\hat{a}_{LX} - \hat{a}_{KX}) - \left[\lambda_{LX}\frac{\lambda_{KY}}{\lambda_{KX}}\theta_{LY} + \lambda_{LY}\theta_{KY}\right](\hat{a}_{KY} - \hat{a}_{LY}) \\ &\quad - \left[\lambda_{LX}\frac{\lambda_{KG}}{\lambda_{KX}}\theta_{LG} + \lambda_{LG}\theta_{KG}\right](\hat{a}_{KG} - \hat{a}_{LG}) - \left[\frac{|\lambda|_{XY} + |\lambda|_{XG}}{\lambda_{KX}}\right]\widehat{Y} = \frac{\varepsilon_G(\theta_{KY} - \theta_{KG})|\lambda|_{XG}}{|\theta|_{XY}\lambda_{KX}}\widehat{T} \\ \text{i.e.,}\quad & - \left[\lambda_{LX}\sigma_X + \left(\lambda_{LX}\frac{\lambda_{KY}}{\lambda_{KX}}\theta_{LY} + \lambda_{LY}\theta_{KY}\right)\sigma_Y + \left(\lambda_{LX}\frac{\lambda_{KG}}{\lambda_{KX}}\theta_{LG} + \lambda_{LG}\theta_{KG}\right)\sigma_G\right](\widehat{W} - \hat{r}) \\ &\quad - \left[\frac{|\lambda|_{XY} + |\lambda|_{XG}}{\lambda_{KX}}\right]\widehat{Y} = \frac{(\theta_{KG} - \theta_{KY})|\lambda|_{XG}}{\lambda_{KX}|\theta|_{XY}}\widehat{T}\end{aligned} \tag{33}$$

Finally, denoting $\eta \equiv \lambda_{LX}\sigma_X + \left(\lambda_{LX}\frac{\lambda_{KY}}{\lambda_{KX}}\theta_{LY} + \lambda_{LY}\theta_{KY}\right)\sigma_Y + \left(\lambda_{LX}\frac{\lambda_{KG}}{\lambda_{KX}}\theta_{LG} + \lambda_{LG}\theta_{KG}\right)\sigma_G$ and substituting values from (9) in the text in (33) yields the solution of $\widehat{Y}$:

$$\widehat{Y} = -\left[\frac{\eta\lambda_{KX} - |\lambda|_{XG}(\theta_{KY} - \theta_{KG})\varepsilon_G}{(\lambda_{KX} - \lambda_{LX})|\theta|_{XY}}\right]\widehat{T}$$

Proceeding as before, from (30) we get,

$$\begin{aligned}\widehat{X} =& \theta_{LX}(\hat{a}_{LX} - \hat{a}_{KX}) - \frac{\lambda_{KY}}{\lambda_{KX}}\theta_{LY}(\hat{a}_{KY} - \hat{a}_{LY}) - \frac{\lambda_{KG}}{\lambda_{KX}}\theta_{LG}(\hat{a}_{GY} - \hat{a}_{LG}) \\ & - \delta\widehat{Y} + \frac{\lambda_{KG}(\theta_{KG} - \theta_{KY})}{\lambda_{KX}|\theta|_{XY}}\widehat{T} \\ =& - \phi(\widehat{W} - \hat{r}) - \delta\widehat{Y} + \frac{\lambda_{KG}(\theta_{KG} - \theta_{KY})}{\lambda_{KX}|\theta|_{XY}}\widehat{T}\end{aligned}$$

which after substitution of values from (9) and (12) in the text boils down to the solution of $\widehat{X}$.

2. The condition for $\varepsilon_G^* > 1$.

By (22) in the text, the critical value of the price elasticity of demand is,

$$\varepsilon_G^* = \frac{(\theta_X + \theta_Y)(\theta_{KX} - \theta_{KG})\lambda_{KX}\delta + \theta_X\phi\lambda_{KX}\delta + \theta_X\gamma}{(\theta_{KY} - \theta_{KG})[\theta_Y(\lambda_{KX} - \lambda_{LX}) - \theta_X(\lambda_{KY} - \lambda_{LY})}$$

Subtracting one from each side we get,

$$\begin{aligned}
\varepsilon_G^* - 1 &= \frac{\theta_X(\theta_{KX} - \theta_{KG})(\lambda_{KX} - \lambda_{LX}) + \theta_X\phi\lambda_{KX}\delta + \theta_X\gamma + \theta_X(\theta_{KY} - \theta_{KG})(\lambda_{KY} - \lambda_{LY}) + \theta_Y(\theta_{KX} - \theta_{KG} - \theta_{KY} + \theta_{KG})(\lambda_{KX} - \lambda_{LX})}{(\theta_{KY} - \theta_{KG})[\theta_Y(\lambda_{KX} - \lambda_{LX}) - \theta_X(\lambda_{KY} - \lambda_{LY})]} \\
&= \frac{\lambda_{LX}\delta[\theta_X\phi + \{(\theta_X + \theta_Y)\theta_{KX} - \theta_X\theta_{KG} - \theta_Y\theta_{KY}\}] + \theta_X(\theta_{KY} - \theta_{KG})(\lambda_{KY} - \lambda_{LY}) + \theta_X\gamma}{(\theta_{KY} - \theta_{KG})[\theta_Y(\lambda_{KX} - \lambda_{LX}) - \theta_X(\lambda_{KY} - \lambda_{LY})]}
\end{aligned}$$

Therefore, a *sufficient* condition for $\varepsilon_G^* > 1$ is that,

$$\theta_{KX} > \frac{\theta_X}{\theta_X + \theta_Y}\theta_{KG} + \frac{\theta_Y}{\theta_X + \theta_Y}\theta_{KY} \equiv \theta_{KX}^*$$

as stated in (23) in the text.

3. Price and Output changes under productivity effect.

Using $\hat{a}_{LX} = -\eta_{LX}(\widehat{W} - \hat{r}) - e_{LX}\widehat{G}$ as in (24) and similar expressions for other input coefficients, (31) now boils down to,

$$\begin{aligned}
&-\left[\lambda_{LX}(\eta_{LX} + \eta_{KX}) + \lambda_{LX}\frac{\lambda_{KY}}{\lambda_{KX}}\eta_{KY} + \lambda_{LY}\eta_{LY} + \lambda_{LX}\frac{\lambda_{KG}}{\lambda_{KX}}\eta_{KG} + \lambda_{LG}\eta_{LG}\right](\widehat{W} - \hat{r}) \\
&\quad -\left[\lambda_{LX}(e_L - e_K) - \lambda_{LX}\frac{\lambda_{KY}}{\lambda_{KX}}e_K + \lambda_{LY}e_L\right]\widehat{G} - \left[\frac{|\lambda|_{XG}}{\lambda_{KX}}\right]\widehat{G} - \left[\frac{|\lambda|_{XY}}{\lambda_{KX}}\right]\widehat{Y} = 0 \\
\Rightarrow \quad &-\eta(\widehat{W} - \hat{r}) - \left[\psi + \frac{|\lambda|_{XG}}{\lambda_{KX}}\right]\widehat{G} - \left[\frac{|\lambda|_{XY}}{\lambda_{KX}}\right]\widehat{Y} = 0
\end{aligned} \tag{34}$$

On the other hand, from the zero profit conditions we get,

$$\begin{aligned}
\widehat{P}_X^* &= 0 = \theta_{LX}\widehat{W} + \theta_{KX}\hat{r} + [\theta_{LX}e_L + \theta_{KX}e_K]\widehat{G} \\
\widehat{P}_Y &= \widehat{T} = \theta_{LY}\widehat{W} + \theta_{KY}\hat{r} + [\theta_{LY}e_L + \theta_{KY}e_K]\widehat{G}
\end{aligned}$$

From these two easy to solve the changes in factor prices as,

$$\widehat{W} = e_L\widehat{G} - \frac{\theta_{KX}}{|\theta|_{KX}}\widehat{T},\ \hat{r} = e_K\widehat{G} - \frac{\theta_{LX}}{|\theta|_{KX}}\widehat{T} \tag{35}$$

$$\Rightarrow \quad \widehat{W} - \hat{r} = (e_L - e_K)\widehat{G} - \frac{1}{|\theta|_{XY}}\widehat{T} \quad [\text{since } \theta_{LX} + \theta_{KX} = 1] \tag{36}$$

which is Eq. (25) in the text.

Then from the zero profit condition for good Z we obtain,

$$\widehat{P}_G = \theta_{LG}\widehat{W} + \theta_{KG}\hat{r}$$

which upon substitution of values from (35) boils down to,

$$\begin{aligned}\widehat{P}_G &= (\theta_{LG}e_L + \theta_{KG}e_K)\widehat{G} - \frac{\theta_{LG}\theta_{KX} - \theta_{KG}\theta_{LX}}{|\theta|_{XY}}\widehat{T} \\ &= (\theta_{LG}e_L + \theta_{KG}e_K)\widehat{G} - \frac{\theta_{KX} - \theta_{KG}}{|\theta|_{XY}}\widehat{T}\end{aligned} \tag{37}$$

which is Eq. (26) in the text.

From the market-clearing condition, on the other hand, we get,

$$-\varepsilon_G(\widehat{P}_G - \widehat{P}_Y) = \widehat{G} - \widehat{Y}$$

which upon substitution of values from (37) reduces to,

$$\begin{aligned}&-\varepsilon_G(\theta_{LG}e_L + \theta_{KG}e_K)\widehat{G} + \frac{\varepsilon_G(\theta_{KX} - \theta_{KG})}{|\theta|_{XY}}\widehat{T} + \frac{\varepsilon_G[\theta]_{GX}}{|\theta|_{XY}}\widehat{T} = \widehat{G} - \widehat{Y} \\ &\widehat{G} = \frac{1}{\phi}\widehat{Y} + \frac{\varepsilon_G(\theta_{KY} - \theta_{KG})}{\phi|\theta|_{XY}}\widehat{T}\end{aligned} \tag{38}$$

where, $\phi \equiv 1 + \varepsilon_G(\theta_{LG}e_L + \theta_{KG}e_K) > 1$.

Substitution of (36) and (38) in (34) yields,

$$\begin{aligned}&-\eta\left[(e_L - e_K)\widehat{G} - \frac{1}{|\theta|_{XY}}\widehat{T}\right] - \left[\psi + \frac{|\lambda|_{XG}}{\lambda_{KX}}\right]\left[\frac{1}{\phi}\widehat{Y} + \frac{\varepsilon_G(\theta_{KY} - \theta_{KG})}{\phi|\theta|_{XY}}\widehat{T}\right] - \left[\frac{|\lambda|_{XY}}{\lambda_{KX}}\right]\widehat{Y} = 0 \\ &-\left[\frac{\eta(e_L - e_K)}{\phi} + \frac{\psi}{\phi} + \frac{|\lambda|_{XG}}{\phi\lambda_{KX}}\frac{|\lambda|_{XY}}{\lambda_{KX}}\right]\widehat{Y} = \left[-\frac{\eta}{|\theta|_{XY}} + \frac{\varepsilon_G(\theta_{KY} - \theta_{KG})\left\{\eta(e_L - e_K) + \psi + \frac{|\lambda|_{XG}}{\lambda_{KX}}\right\}}{\phi|\theta|_{XY}\lambda_{KX}}\right]\widehat{T} \\ &\Rightarrow \quad \widehat{Y} = -\frac{\left[-\eta\phi + \varepsilon_G(\theta_{KY} - \theta_{KG})\left\{\eta(e_L - e_K) + \psi + \frac{|\lambda|_{XG}}{\lambda_{KX}}\right\}\right]}{\left[\frac{\eta(e_L - e_K)}{\phi} + \frac{\psi}{\phi} + \frac{|\lambda|_{XG}}{\phi\lambda_{KX}}\frac{|\lambda|_{XY}}{\lambda_{KX}}\right]|\theta|_{XY}}\widehat{T}\end{aligned} \tag{39}$$

Substitution of (39) in (38) then yields,

$$
\begin{aligned}
&\widehat{G} = -\frac{\left[-\eta\phi + \varepsilon_G(\theta_{KY} - \theta_{KG})\frac{|\lambda|_{XY}}{\lambda_{KX}}\right]}{\left[\eta(e_L - e_K) + \psi + \frac{|\lambda|_{XG}}{\lambda_{KX}} + \frac{\phi|\lambda|_{XY}}{\lambda_{KX}}\right]|\theta|_{XY}}\widehat{T} \\
\Rightarrow\quad &\widehat{G} = -\frac{\left[-\eta\phi\lambda_{KX} + \varepsilon_G(\theta_{KY} - \theta_{KG})|\lambda|_{XY}\right]}{\left[\eta(e_L - e_K)\lambda_{KX} + \psi\lambda_{KX} + |\lambda|_{XG} + \phi|\lambda|_{XY}\right]|\theta|_{XY}}\widehat{T} \\
\Rightarrow\quad &\widehat{G} = -\frac{\left[-\eta\phi\lambda_{KX} + \varepsilon_G(\theta_{KY} - \theta_{KG})|\lambda|_{XY}\right]}{\left[|\lambda|_{XG} + |\lambda|_{XY}\right]|\theta|_{XY} + \left[\eta(e_L - e_K)\lambda_{KX} + \psi\lambda_{KX} + +(\phi - 1)|\lambda|_{XY}\right]|\theta|_{XY}}\widehat{T} \\
&\widehat{G} = -\left[\frac{\eta\lambda_{KX} + (\theta_{KY} - \theta_{KG})|\lambda|_{XY}\varepsilon_G}{(\lambda_{KX} - \lambda_{LX})|\theta|_{XY} - \left\{\eta(e_L - e_K)\lambda_{KX} + \psi\lambda_{KX} + (\phi - 1)|\lambda|_{XY}\right\}|\theta|_{XY}}\right]\widehat{T}
\end{aligned}
\tag{40}
$$

which is as specified in (27).

Finally, to check the sign of ψ, note that by definition,

$$
\begin{aligned}
\psi &= (\lambda_{LX} + \lambda_{LY})e_L - \frac{(\lambda_{KX} + \lambda_{KY})\lambda_{LX}}{\lambda_{KX}}e_K \\
&= (\lambda_{LX} + \lambda_{LY})\left[e_L - \frac{(\lambda_{KX} + \lambda_{KY})\lambda_{LX}}{(\lambda_{LX} + \lambda_{LY})\lambda_{KX}}e_K\right]
\end{aligned}
\tag{41}
$$

Now, $\frac{(\lambda_{KX} + \lambda_{KY})\lambda_{LX}}{(\lambda_{LX} + \lambda_{LY})\lambda_{KX}} - 1 = \frac{\lambda_{KX}\lambda_{LX} + \lambda_{KY}\lambda_{LX} - \lambda_{KX}\lambda_{LX} - \lambda_{KX}\lambda_{LY}}{(\lambda_{LX} + \lambda_{LY})\lambda_{KX}}$

$$
= \frac{\lambda_{KY}\lambda_{LX} - \lambda_{KX}\lambda_{LY}}{(\lambda_{LX} + \lambda_{LY})\lambda_{KX}}
$$

which is positive by the factor intensity assumption. That is, the coefficient of e_K in (41) is larger than unity so that $\psi < 0$ for $e_L \le e_K$. For $e_L > e_K$ also ψ may be negative if e_L is not too large.

References

Alesina A, Wacziarg R (1998) Openness, country size and government. J Public Econ 69(3):305–321

Annet A (2001) Social fractionalization, political instability and the size of the government. IMF Staff Papers 48(3):561–592

Anwar S, Zheng M (2004) International factor mobility, government size and welfare. Australian Economic Papers, June, pp 248–255

Blomberg SB (1996) Growth, political instability and the defense burden. Economica 63:649–672

Rodrik D (1998) Why do more open economies have bigger governments? J Polit Econ 106:997–1032

Velasco A (1997) A model of endogenous fiscal deficits and delayed fiscal reforms. NBER working paper # 6336

The Effects of Corruption on Trade Flows: A Disaggregated Analysis

Subhayu Bandyopadhyay and Suryadipta Roy

Abstract This paper analyzes the effects of corruption on trade flows for a panel of nations. The novelty of the paper lies in our focus on trade flows at the industry level. Such a focus is important because the emerging literature on trade and institutions suggests that the pattern of comparative advantage between nations may be driven by international differences in institutions, among other factors. We find that exports of certain goods by the developed nations are negatively affected by higher domestic corruption in these nations. Interestingly, corruption in the trading partner also reduces exports of certain goods. Imports seem somewhat less vulnerable to corruption. The analysis uncovers considerable heterogeneity in the effects of corruption on different industries.

1 Introduction

It is widely agreed that institutional quality matters greatly for international trade. Strong and supportive institutions are supposed to reduce the uncertainty involved in international transactions, and thereby reduce the transactions cost of trade. On the contrary, high levels of bureaucratic corruption, weak rule of law, or lack of

The views expressed are those of the authors and do not necessarily represent official positions of the Federal Reserve Bank of St. Louis or of the Federal Reserve System.

S. Bandyopadhyay (✉)
Federal Reserve Bank of St. Louis, Research Division, PO Box 442, St. Louis, MO 63166-0442, USA
e-mail: Subhayu.Bandyopadhyay@stls.frb.org

S. Bandyopadhyay
IZA, Bonn, Germany

S. Roy
Department of Economics, Earl N. Phillips School of Business, High Point University, High Point,, NC 27262, USA
e-mail: sroy@highpoint.edu

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_6

contract enforcement increase the uncertainty about the size of expected gains from carrying out an international transaction, and thereby deter international trade. Nunn and Trefler (2014) have surveyed the theoretical literature and empirical evidences on the role of domestic institutional quality of countries in affecting their comparative advantage. While trade theorists have traditionally focused on the role of factor endowments and technological differences between countries in predicting trade patterns, it has been observed that trade volumes are much lower than the levels predicted by the theoretical models, and high income capital-abundant countries trade disproportionately with each other rather than with the low income labor-abundant countries (see Trefler 1995). While this case of "missing trade" led to the emergence of the so-called New Trade Theories (N-T) developed by Helpman and Krugman (1985) that emphasized the role of increasing returns to scale and imperfect competition, the literature has recently focused on institutional differences between countries as a major explanatory factor of bilateral trade patterns that was not adequately studied by the erstwhile trade theories. The idea is that international transactions occur in the context of asymmetric information, opportunistic behavior, and transactions costs that in turn are governed by the institutional quality of countries (see Belloc 2006).

Anderson and Marcouiller (2002) view institutional quality reflected in corruption and imperfect contract enforcement as indicators of transactions costs of trade. Using the gravity model of trade, they find that poor institutions affect trade as much as formal trade barriers like tariffs. Ranjan and Lee (2007) investigate a particular aspect of institutions—contract enforcement, and find that their proxy measures of contract enforcement affect the volume of trade in both differentiated and homogeneous goods. de Groot et al. (2004) have investigated the effect of various measures of institutional quality from the World Governance Indicators, and find better institutional quality to lead to higher bilateral trade. In similar lines, Mohlmann et al. (2010) view institutional differences between countries as measures of intangible trade barriers, and point to heterogeneous effects of such barriers on bilateral trade in homogeneous and heterogeneous goods. Implementing a Poisson estimator on panel data on bilateral trade flows between countries, Francois and Manchin (2013) find that bilateral trade depends on both institutional quality as well as infrastructure. Thus by now, there is a broad consensus on the critical role that institutions play in promoting bilateral trade, and the pernicious effect of poor institutional quality in hampering trade between countries—both at the aggregate level as well as at the sectoral level.

The above studies however, do not take into account the heterogeneous effect of institutions on the level of economic development of countries in affecting bilateral trade. Marjit et al. (2014) have studied this interaction between corruption and economic development in affecting trade openness of countries. Viewing corruption as a labor-intensive activity and by introducing corruption as a source of (negative) factor endowment in the Heckscher–Ohlin–Vanek model of trade, they show that corruption distorts relative factor endowment of countries and thereby has impact on comparative advantage and resulting trade patterns. While their paper empirically investigates if the effect of corruption on aggregate trade openness (measured by Trade-GDP ratio) depends on factor abundance, the current paper investigates the interaction between corruption and economic development on bilateral trade between countries at a disaggregated level using sectoral trade data. The motivation is as follows: given that high income developed countries usually specialize and export more differentiated products that use more capital-intensive production methods in contrast to the underdeveloped countries that specialize more on homogeneous goods using (relatively) more labor-intensive techniques, greater corruption should impact relative factor endowments and have independent effect on bilateral trade beyond its effects on aggregate trade openness. We investigate this hypothesis empirically by implementing the gravity model of trade using bilateral export and import data for seven manufacturing industries according to the 3-digit ISIC Rev. 2 classification that vary based on factor intensities, extent of scale economies, and degree of product differentiation. The major findings can be summarized as follows. First, based on the country pair and year fixed effects specifications, we find strong evidence of negative impact of domestic corruption on the high income exporters, in that greater corruption in these countries is associated with reduced bilateral exports in case of textiles, industrial chemicals, rubber products, and transport equipment industries as well as a modest negative impact in the tobacco industry. Similarly, greater corruption in the trading partners is associated with reduced exports for the high income exporters in the electric machinery, transport equipment, industrial chemicals, and rubber products industry. Second, the negative effects of domestic corruption on the developed countries are more pronounced on exports than imports. Domestic corruption is not found to be associated with any reduction in imports. On the other hand, there is some evidence of positive effect of domestic corruption for the relatively developed countries' imports in case of the transport equipment industry. We also find substantial heterogeneity in the impact of partner country corruption on imports. Greater partner country corruption is associated with higher imports for the relatively advanced countries in the electric machinery and in the transport equipment industry, and lower imports in the textiles and in the footwear industry. Third, the effect of corruption seems to depend on the degree of product differentiation in that the corruption terms as well as their interaction with per capita income are found to be not significant in case of the tobacco industry—arguably the least differentiated and most homogeneous sector in our sample.

The paper is conceptually related to the literature on the role of institutions as a source of comparative advantage as surveyed in Nunn and Trefler (2014). However, while the focus in these papers is mainly on how domestic institutions shape the comparative advantage of countries, the current paper is more interested in investigating the deleterious impact on bilateral trade of one particular aspect of institutional quality—namely the corruption level of countries, in different industries and how the above effect varies systematically between the high income and the low income countries. In its application of the gravity model to explain bilateral trade in different sectors, the paper is methodologically linked to the literature on the gravity model of sectoral trade (see van Bergeijk and Brakman (2010) for a comprehensive survey on the role of the gravity model in international trade).

The rest of the paper is organized as follows. Section 2 presents the gravity model and the different empirical specifications to be estimated. Section 3 discusses the data sources and presents the main results. Section 4 concludes the paper and discusses some avenues for further research.

2 Empirical Model

The gravity model has long been a workhorse for analyzing the determinants of bilateral trade flows. It has been extensively applied to analyze effects of policy-related as well as behind-the-border trade barriers, regional trading groups, and economic integration policies. The model also owes its popularity to a large extent due to its predictive power to explain bilateral trade flows as measured by the goodness of fit. In its most intuitive version, the gravity model postulates that bilateral trade depends on the economic size of the trading partners, which reflects market size and purchasing power, and a measure of economic distance (or proximity) between countries to reflect trade costs. Some of the early theoretical foundations for the gravity model have been provided by Anderson (1979) using the Armington assumption whereby goods are differentiated by their country of origin and where consumers have preferences over all the differentiated products. Subsequently, Bergstrand (1985, 1989) have shown that the gravity model can be derived from trade models based on monopolistic competition where identical countries trade differentiated goods because consumers have a taste for variety. Deardorff (1998) has derived gravity-type expressions from the standard Heckscher–Ohlin trade model, and has shown that the gravity equation can be consistent with a large class of trade models. More recently, in a significant development, Anderson and van Wincoop (2003) have shown that controlling for relative trade costs is crucial for well-specified gravity models. Their theoretical results suggest that bilateral trade does not only depend on bilateral trade cost between two countries, but also on the relative weight of these costs with respect to the other trade partners (what they refer to as multilateral resistance). Thus, it is important to account for the multilateral resistance terms in the estimation of gravity models.

Given that we focus on the asymmetric impact of corruption between the high income and the low income countries toward their bilateral exports and imports in different industries, corruption in our model is viewed as a component of trade costs. Thus, we augment the traditional gravity model to include a number of country-specific and dyadic factors as components of trade costs. Besides the exporter-and importer-level corruption terms (Corr_{it} and Corr_{jt}), these include distance between the exporter and the importer country, whether the country pair belongs to a regional trade agreement or the existence of any preferential trading agreement, or whether the trade partners share a common language, border or colonial heritage with its trade partner. Our baseline log-linear gravity specification to be estimated for exports of sector k products from country i into country j in year t (X_{ijt}^{k}) is as follows:

$$\ln X_{ijt}^{k} = \beta_{1}^{k} \ln Y_{it} + \beta_{2}^{k} \ln Y_{jt} + \beta_{3}^{k} |\ln Y_{it} - \ln Y_{jt}| + \delta_{1}^{k} \ln y_{it} + \delta_{2}^{k} \ln y_{jt} + \delta_{3}^{k} |\ln y_{it} - \ln y_{jt}| + \gamma_{1}^{k} \mathrm{Corr}_{it} + \gamma_{2}^{k} \mathrm{Corr}_{jt} + \gamma_{12}^{k} \mathrm{Corr}_{it} * \ln y_{it} + \gamma_{21}^{k} \mathrm{Corr}_{jt} * \ln y_{it} + \theta'^{k} Z_{ijt} + a_{ij}^{k} + a_{t}^{k}. \tag{1}$$

The above empirical specification is a reduced form equation that reflects both increasing returns to scale with monopolistic competition, and the factor proportions theory of trade. Export flows are expected to be directly related to the market size of the trading partners reflected by the real gross domestic product (GDP) (Y_{it}/Y_{jt}). Relative factor endowments measured by real per capita GDP (y_{it}/y_{jt}) are also important determinants of trade flows. According to Markusen and Venables (2000), a large part of manufacturing trade is in the form of intra-industry trade between multinational firms from developed countries. More specifically, with positive trade costs, multinationals are more likely to exist when the countries are more similar in both relative and absolute endowments, and thereby substitute bilateral trade by horizontal foreign direct investment (FDI). Hence, we introduce the logarithm of the absolute difference in the GDP of country i and j to reflect the substitution away from bilateral trade toward FDI. We also introduce the logarithm of the absolute difference in the GDP per capita between the trading partners. This variable controls for factor endowment differences between countries, and can be treated as a proxy for wage differences between countries. a_{ij}-s are fixed effects capturing time-invariant country-pair-specific heterogeneity that control for multilateral resistance in the gravity model, and the a_t-s are year fixed effects capturing time-varying heterogeneity. Z_{ijt} denotes the vector of pair-specific dummy variables referred to above that affect trade costs between countries. The baseline specification implements a pair (i.e., exporter–importer) fixed effects estimator using panel data for different industries. While this has the advantage of mitigating the bias generated by heterogeneity across countries, the pair fixed effects absorb all heterogeneity that is constant over time, and hence, we cannot recover these estimates included in Z_{ijt} (e.g., distance, common language, etc.). In order to address this problem, we also estimate a random effects model proposed by Mundlak (1978). This model includes averages of the time variant variables to control for the remaining bilateral unobserved heterogeneity that is time invariant and could be correlated with the idiosyncratic error term. This model accounts for bilateral time-invariant heterogeneity that is correlated with the regressors and allows us to estimate the coefficients of the time-invariant variables.

The key terms in the above equation are the interaction terms between domestic and partner country corruption and per capita GDP, i.e., $\mathrm{Corr}_{it} * \ln y_{it}$ and $\mathrm{Corr}_{jt} * \ln y_{it}$ respectively. From Eq. 1,

$$\frac{\partial \ln X_{ijt}^{k}}{\partial \mathrm{Corr}_{it}} = \gamma_{1}^{k} + \gamma_{12}^{k} \ln y_{it} \tag{1a}$$

For relatively low income countries, the γ_1^k term dominates, and the effect of domestic corruption on bilateral exports will be determined by the sign of γ_1^k. On the other hand, the effect of domestic corruption on bilateral exports for the high income countries will be determined by γ_{12}^k. Similarly, the effect of partner country corruption on bilateral exports can be obtained as

$$\frac{\partial \ln X_{ijt}^k}{\partial \mathrm{Corr}_{jt}} = \gamma_2^k + \gamma_{21}^k \ln y_{it} \tag{1b}$$

Thus, the specification allows us to investigate if the effect of corruption on bilateral exports in different industries depends on the level of economic development proxied by GDP per capita. Given that advanced economies possess better quality institutions reflected in lower levels of corruption, stronger property rights, and better contract enforcement, they tend to have comparative advantage in products whose transactions costs are more sensitive to the quality of institutions (e.g., Berkowitz et al. 2006; Nunn 2007, Levchenko 2007; Ranjan and Lee 2007). Empirical evidence based on these papers suggest that the more institutionally intensive industries usually involve higher levels of product differentiation, are more "high-tech" in nature, use relatively higher amounts of capital inputs, and involve greater amount of relationship-specific investment in their production process (Rauch (1999)). Thus greater domestic or partner corruption is more likely to affect exports of institutionally intensive goods from the high income countries compared to the low income countries. Moreover, by controlling for differences in factor endowments between countries, the specification also isolates the independent effect of institutions on comparative advantage and thereby on bilateral exports. A similar specification is also estimated for imports of sector k products by country i from country j in year t (M_{ijt}^k) to understand the difference in the effect of domestic and partner institutions on the high income and the low income countries:

$$\begin{aligned}\ln M_{ijt}^k =& \tilde{\beta}_1^k \ln Y_{it} + \tilde{\beta}_2^k \ln Y_{jt} + \tilde{\beta}_3^k |\ln Y_{it} - \ln Y_{jt}| + \tilde{\delta}_1^k \ln y_{it} + \tilde{\delta}_2^k \ln y_{jt} + \tilde{\delta}_3^k |\ln y_{it} - \ln y_{jt}| \\ &+ \tilde{\gamma}_1^k \mathrm{Corr}_{it} + \tilde{\gamma}_2^k \mathrm{Corr}_{jt} + \tilde{\gamma}_{12}^k \mathrm{Corr}_{it} * \ln y_{it} + \tilde{\gamma}_{21}^k \mathrm{Corr}_{jt} * \ln y_{it} + \tilde{\theta}'^k Z_{ij} + \tilde{a}_{ij}^k + \tilde{a}_t^k.\end{aligned} \tag{2}$$

Given that the high-income countries tend to have comparative advantage on institutionally intensive industries, it is likely that they will import less institutionally intensive goods and the low income countries will import more institutionally intensive products. Hence greater domestic or partner corruption is more likely to affect the imports of institutionally intensive goods by the low income countries compared to the high income countries. However, given that the high income countries enjoy lower corruption compared to the low income countries, these countries are likely to have lower transactions costs and hence experience higher imports. Thus, the net asymmetric effect of corruption on imports between the high income and the low income countries in different sectors is more of an empirical question that we take to the data.

3 Data Sources and Results

Disaggregated data on bilateral exports and imports have been obtained from Nicita and Olarreaga (2006). Their database provides trade, production, and protection data for 28 manufacturing sectors at the 3-digit level International Standard Industrial Classification (ISIC) Revision 2 level for potentially 100 developing and developed countries over the period 1976–2004. The current study uses the following industries: tobacco (314), textiles (321), footwear (324), industrial chemicals (351), rubber products (355), electrical machinery (383), and transport equipment (384). Real GDP and real GDP per capita (both in US dollars) have been obtained from Nicita and Olarreaga (2006). Information on country-pair-specific variables like distance between countries i and j, dummy variables indicating presence of regional trading agreement, accordance of preferential trading agreement under the Generalized System of Preferences (GSP) by a developed country to a developing country trade partner, whether they share a common border or common language or have common colonial origin have been obtained from Head et al. (2010). Our measure of corruption is from the International Country Risk Guide (ICRG) published by the Political Risk Services (PRS) Group. This is an assessment of corruption within the political system reflecting demand for special payments and bribes to obtain regular official services (financial corruption) as well as other distortionary practices like government patronage or special favors for select business interests (political corruption). Both forms of corruption distort economic and financial incentives, reduce efficiency, and introduce political instability. The index ranges from 0 to 6 with higher values denoting very low corruption; for interpretation purposes, the index has been rescaled to vary from 0 to 1 such that higher values of $\mathrm{Corr}_{it}/\mathrm{Corr}_{jt}$ denote greater corruption. Corruption data are available for 174 countries during 1984–2010 which is broader than the other commonly used sources of corruption data like the Transparency International or the World Bank's Worldwide Governance Indicators; however, the data coverage is not uniform for all countries. With the exception of the tobacco industry, our baseline fixed effects specification is based on observations for 91 countries over the period 1984–2010.[1] The list of countries used in the regression has been reported in the Appendix 1.

The gravity specification in Eq. 1 attempts to uncover the asymmetric effect of corruption on bilateral trade in different sectors between the high income and the low income countries. We begin by reporting the results for the random effects model that allows us to estimate the coefficients for the time-invariant determinants of trade costs in gravity models. The results for the seven manufacturing industries are reported in Table 1. Exports are expected to be directly related to the market size of the trade partners measured by real GDP. Feenstra et al. (2001) have derived gravity-type equations for bilateral trade in differentiated and homogeneous goods,

[1]For the tobacco industry, bilateral trade data have been reported for 87 exporting countries and 91 importing countries.

and suggest the heterogeneous goods to be characterized by home market effect with larger export elasticity with respect to the exporter's income than with respect to the importer's income (Krugman (1980)) while the homogeneous goods have a reversed home market effect. They also find strong evidence to this hypothesis. Based on the results in Table 1, all industries except tobacco share the characteristics of heterogeneous goods with the effect of the home country GDP being greater than the partner country's GDP. In case of the transport equipment industry, exports are found to be strongly negatively related to the importer's GDP which suggests strong home market bias in this industry. On the other hand, based on the Feenstra et al. (2001) framework, the tobacco industry can be characterized as homogeneous given that importer's elasticity is greater than the exporter's elasticity, and the latter being not significant. Markusen and Venables (2000) have suggested that two countries similar in economic size are likely to experience greater multinational activity, e.g., through foreign direct investment instead of bilateral trade. Based on Table 1, we observe that relative country size is positively associated with bilateral exports in the rubber products industry and reduced bilateral trade in the footwear industry. Thus, it is likely that bilateral trade has been substituted by foreign direct investment in the footwear industry. Moreover, it is also possible that unilateral trade preference granted by the developed countries to their developing trade partners under the GSP program have promoted trade between countries that are dissimilar in economic size as is indicated by the statistically significant GSP dummy in case of the rubber products industry. The only industry where bilateral trade seems to be driven by difference in factor endowments according to the Heckscher–Ohlin framework is the tobacco industry, where bilateral trade is positively related to the difference in per capita GDP between the trade partners. The results for the other variables mostly conform to the existing literature on the gravity model. Bilateral exports in all industries are negatively related to the distance between the trade partners, and positively related to the dummy variables that indicate whether the trading partners share a common language, religion, and colonial past (Mohlmann et al. (2010)). The only exception is that the common language dummy is found to be not significant in case of the tobacco industry. This finding is consistent with the network/search theory of Rauch (1999) which suggests that search costs (e.g., those affected by shared language, shared links or geographical proximity) are likely to be least important for homogeneous goods.

Moving on to the interaction terms between corruption and GDP per capita, we find strong evidence of asymmetric effect of domestic corruption between the high income and the low income countries in the textiles, industrial chemicals, rubber products, and in the transport equipment industry. Domestic corruption appears to reduce bilateral exports in the more advanced countries and increase exports in the less developed countries. We also find similar asymmetric effect of partner country corruption in the industrial chemicals, rubber products, electric machinery, and in the transport equipment industry. There does not appear to be any difference in the effect of corruption between countries in case of the footwear industry and in the tobacco industry, although we observe a moderate negative impact of domestic

Table 1 Correlated random effects model results. Dependent variable: ln(bilateral exports)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ISIC classification	Tobacco (314)	Textiles (321)	Footwear (324)	Industrial chemicals (351)	Rubber products (355)	Electric machinery (383)	Transport equipment (384)
ln(RealGDP(*i*))	0.38	2.26***	1.41***	3.17***	2.02***	1.70***	0.51*
	[0.64]	[9.95]	[4.28]	[15.50]	[8.14]	[7.59]	[2.05]
ln(RealGDP(*j*))	1.28**	1.44***	1.02***	0.39*	−0.25	−0.23	−0.82***
	[3.19]	[6.95]	[3.80]	[2.35]	[1.33]	[1.22]	[3.96]
ln|(RealGDP (*i*) − RealGDP(*j*))|	−0.08	0.04	−0.10*	−0.01	0.10*	0.05	0.05
	[0.97]	[1.07]	[1.99]	[0.42]	[2.49]	[1.58]	[1.21]
ln(GDPpercapita(*i*))	0.79	−0.65**	−0.28	−0.74***	−0.68*	0.44^{+}	1.60***
	[1.25]	[2.82]	[0.77]	[3.37]	[2.35]	[1.88]	[5.96]
ln(GDPpercapita(*j*))	−0.46	−0.47*	0.32	0.43*	1.26***	1.69***	1.77***
	[1.05]	[2.33]	[1.15]	[2.55]	[6.70]	[9.29]	[8.62]
ln|(GDPpercapita (*i*) − GDPpercapita(*j*))|	0.07*	−0.02	-0.04^{+}	−0.03	0.00	−0.02	−0.00
	[1.99]	[0.86]	[1.85]	[1.46]	[0.17]	[1.04]	[0.15]
RTA	0.07	0.10^{+}	−0.01	0.04	−0.02	−0.06	−0.01
	[0.64]	[1.92]	[0.15]	[0.89]	[0.45]	[1.30]	[0.09]
GSP	−0.27	0.11	−0.05	-0.27^{+}	0.38*	−0.03	−0.10
	[0.59]	[0.67]	[0.18]	[1.87]	[2.40]	[0.18]	[0.62]
Corruption(*i*)	0.73	0.63***	0.01	0.45*	1.25***	−0.07	0.38^{+}
	[1.64]	[3.58]	[0.04]	[2.50]	[5.83]	[0.34]	[1.74]
Corruption(*j*)	0.37	−0.19	−0.12	0.43*	0.66**	1.07***	0.83***
	[0.82]	[1.15]	[0.47]	[2.56]	[3.08]	[5.55]	[4.48]
Corruption(*i*) * ln (GDPpercapita(*i*))	-0.09^{+}	−0.07***	0.00	−0.05**	−0.14***	0.01	−0.05*
	[1.83]	[3.84]	[0.15]	[2.78]	[6.00]	[0.48]	[2.21]

(continued)

Table 1 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ISIC classification	Tobacco (314)	Textiles (321)	Footwear (324)	Industrial chemicals (351)	Rubber products (355)	Electric machinery (383)	Transport equipment (384)
Corruption(*j*) * ln (GDPpercapita(*i*))	−0.05	0.02	0.00	−0.04*	−0.06**	−0.11***	−0.09***
	[1.09]	[1.08]	[0.06]	[2.33]	[2.85]	[5.43]	[4.41]
ln(distance)	−0.97***	-1.08***	−0.88***	-1.27***	−0.98***	-1.11***	-1.09***
	[15.35]	[26.85]	[17.44]	[34.07]	[23.26]	[24.94]	[25.51]
Common border	0.51*	0.91***	0.91***	1.01***	0.95***	0.93***	1.09***
	[2.56]	[6.73]	[5.32]	[8.48]	[6.52]	[6.09]	[7.08]
Colonial relationship	0.84***	1.01***	0.66***	0.65***	0.66***	0.91***	1.14***
	[4.27]	[7.03]	[3.63]	[4.65]	[4.41]	[5.49]	[7.29]
Common language	0.19	0.31***	0.24**	0.57***	0.21**	0.61***	0.41***
	[1.64]	[4.13]	[2.63]	[7.65]	[2.77]	[7.37]	[5.56]
Constant	-2.35^{+}	-24.95***	-18.14***	-29.94***	-20.33***	-33.62***	-31.18***
	[1.83]	[34.29]	[17.67]	[40.02]	[24.69]	[42.94]	[40.78]
N	23785	74935	42843	72664	58284	73015	66832
RMSE	1.59	1.19	1.30	1.15	1.14	1.24	1.47
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Chi-squared	62.53***	160.56***	98.36***	414.27***	550.26***	524.46***	336.87***

Note $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; Absolute t statistics in brackets; Higher values of the corruption variable denote greater corruption. All regressions based on robust standard errors clustered within country pairs corruption in the tobacco industry. Overall these results suggest asymmetric effect of corruption in the more heterogeneous industries. On the other hand, in a more homogeneous industry (i.e., tobacco) or an industry where bilateral trade has been substituted by multinational activity (e.g., footwear), there is little or no evidence of such asymmetric effect of corruption. While the results are interesting, the chi-square tests following the random effects model suggest that the assumption of fixed effects being uncorrelated with the explanatory variables in the random effects model might be violated. Hence, we implement the fixed effects estimator using Eq. 1, and the results are reported in Table 2. Based on the goodness of fit, the gravity models are able to explain substantial amount of variability in bilateral trade. While the individual estimates of the variables for the fixed effects model are different from the random effects model, we obtain similar evidence for the asymmetric effect of domestic and partner corruption on bilateral exports in the industrial chemicals, rubber products, electric machinery, and in the transport equipment industry.

In order to investigate if the various determinants of trade and the effects of corruption differ with respect to their impact on the directions of trade flows, i.e., exports and imports, we estimate Eq. 2 for bilateral imports using the random effects model and the fixed effects model. Since we are more interested in the estimates of the time-varying determinants of bilateral trade, and the assumption of independence of the fixed effects with the explanatory variables in the random effects model are found to be violated, here we only report the results for the fixed effects model.[2] The results for the log-linear gravity specification for bilateral imports of sector k products by importer i from exporter j are reported in Table 3. The model for bilateral imports suggests home market effect in the industrial chemicals, rubber products, electric machinery, and in the transport equipment industry similar to the model for bilateral exports, given that income elasticity of imports for the exporter country is found to be greater than that for the importer country for these sectors. Also, bilateral imports are found to be positively associated with relative country size in the rubber products industry.

The effect of per capita incomes on bilateral imports is a topic of great interest in the literature. Anderson and Marcouiller (2002) find that the share of total expenditure on traded goods declines with increase in per capita income. This is possible if countries with higher incomes reduce the share of expenditure on traded goods and increase their share of expenditure toward nontraded goods and services. Markusen (2013) has recently emphasized that introducing nonhomothetic preference in the traditional trade models can help explain the significance of per capita income on bilateral trade in the gravity equations. Our results suggest that nonidentical and nonhomothetic preferences play a role in explaining imports in the

[2]The effects of the time-invariant variables on bilateral imports in the random effects model are found to be exactly similar to that for bilateral exports. Imports are found to be positively associated with sharing a common border, having a common colonial heritage, and sharing a common language, while negatively associated with distance between the trading partners. The results can be obtained from the authors on request.

Table 2 Linear fixed effects model results. Dependent variable: ln(bilateral exports)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ISIC classification	Tobacco (314)	Textiles (321)	Footwear (324)	Industrial chemicals (351)	Rubber products (355)	Electric machinery (383)	Transport equipment (384)
ln(RealGDP(*i*))	0.49	2.29***	1.22***	3.20***	2.08***	1.73***	0.53*
	[0.71]	[9.42]	[3.42]	[14.66]	[7.86]	[7.28]	[1.96]
ln(RealGDP(*j*))	1.37**	1.37***	0.88**	0.30^{+}	−0.42*	−0.33	-1.02***
	[3.06]	[6.22]	[3.03]	[1.73]	[2.06]	[1.64]	[4.53]
ln|(RealGDP (*i*) − RealGDP(*j*))|	−0.08	0.04	-0.10^{+}	−0.01	0.11*	0.05	0.05
	[0.90]	[1.04]	[1.83]	[0.38]	[2.50]	[1.55]	[1.21]
ln(GDPpercapita(*i*))	0.69	−0.74**	−0.19	−0.80***	−0.83**	0.34	1.50***
	[0.98]	[3.01]	[0.49]	[3.43]	[2.70]	[1.37]	[5.23]
ln(GDPpercapita(*j*))	−0.52	−0.43*	0.41	0.50**	1.41***	1.77***	1.92***
	[1.10]	[2.00]	[1.36]	[2.81]	[7.02]	[9.26]	[8.83]
ln|(GDPpercapita(*i*) −GDPpercapita(*j*))|	0.07^{+}	−0.02	-0.05^{+}	−0.03	0.00	−0.02	−0.00
	[1.83]	[0.83]	[1.78]	[1.41]	[0.15]	[0.99]	[0.04]
RTA	0.08	0.10^{+}	−0.01	0.04	−0.03	−0.06	−0.01
	[0.72]	[1.89]	[0.12]	[0.87]	[0.58]	[1.22]	[0.10]
GSP	−0.35	0.11	−0.06	−0.23	0.37*	−0.04	−0.11
	[0.68]	[0.62]	[0.21]	[1.52]	[2.22]	[0.24]	[0.63]
Corruption(*i*)	0.72	0.66***	0.14	0.50**	1.36***	−0.03	0.48*
	[1.49]	[3.61]	[0.45]	[2.61]	[5.93]	[0.13]	[2.07]
Corruption(*j*)	0.49	−0.28	−0.22	0.45*	0.53*	1.10***	0.81***
	[0.87]	[1.50]	[0.74]	[2.37]	[2.20]	[5.02]	[3.69]
Corruption(*i*) * ln (GDPpercapita(*i*))	-0.09^{+}	−0.08***	−0.01	−0.06**	−0.15***	0.01	−0.06*
	[1.67]	[3.86]	[0.32]	[2.91]	[6.12]	[0.23]	[2.54]

(continued)

Table 2 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ISIC classification	Tobacco (314)	Textiles (321)	Footwear (324)	Industrial chemicals (351)	Rubber products (355)	Electric machinery (383)	Transport equipment (384)
Corruption(*j*) * ln (GDPpercapita(*i*))	−0.06	0.03	0.01	−0.04*	−0.05*	−0.11***	−0.08***
	[1.07]	[1.44]	[0.37]	[2.18]	[2.01]	[4.92]	[3.64]
Constant	-43.19**	-76.32***	-48.05***	-78.78***	-45.97***	-49.61***	-14.28*
	[3.05]	[12.60]	[5.80]	[16.25]	[7.98]	[8.91]	[2.29]
N	23785	74935	42843	72664	58284	73015	66832
R^2	0.754	0.859	0.821	0.872	0.849	0.886	0.837
adj. R^2	0.715	0.846	0.801	0.860	0.834	0.875	0.821
RMSE	1.58	1.19	1.30	1.15	1.14	1.23	1.47
Country-pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; Absolute t statistics in brackets; Higher values of the corruption variable denote greater corruption. All regressions based on robust standard errors clustered within country pairs

Table 3 Linear fixed effects model results. Dependent variable: ln(bilateral imports)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ISIC classification	Tobacco (314)	Textiles (321)	Footwear (324)	Industrial chemicals (351)	Rubber products (355)	Electric machinery (383)	Transport equipment (384)
ln(RealGDP(*i*))	0.80	1.77***	0.87*	0.45*	−0.49*	0.68**	−0.69**
	[1.50]	[8.02]	[2.54]	[2.33]	[2.22]	[3.17]	[2.77]
ln(RealGDP(*j*))	0.08	1.77***	0.87*	2.68***	2.34***	1.49***	0.71**
	[0.12]	[8.03]	[2.41]	[13.63]	[9.12]	[7.11]	[2.74]
ln|(RealGDP (*i*) − RealGDP(*j*))|	−0.10	0.00	−0.02	−0.04	0.12**	0.06^{+}	−0.02
	[1.05]	[0.10]	[0.36]	[1.35]	[3.22]	[1.78]	[0.37]
ln(GDPpercapita(*i*))	−0.25	−0.82***	0.53	0.26	1.33***	0.81***	1.17***
	[0.47]	[3.75]	[1.55]	[1.44]	[6.10]	[3.94]	[4.82]
ln(GDPpercapita(*j*))	0.79	-0.40^{+}	0.37	−0.76***	−0.91**	0.45*	1.57***
	[1.10]	[1.87]	[0.99]	[3.69]	[3.06]	[2.07]	[5.67]
ln|(GDPpercapita (*i*) − GDPpercapita(*j*))|	0.07^{+}	0.02	0.01	−0.00	0.03	−0.02	0.01
	[1.66]	[0.92]	[0.31]	[0.28]	[1.43]	[1.30]	[0.47]
RTA	0.07	0.04	0.03	0.04	0.04	−0.06	0.01
	[0.51]	[0.78]	[0.37]	[1.06]	[0.64]	[1.29]	[0.12]
GSP	−0.15	−0.02	−0.20	0.21*	0.24^{+}	−0.12	−0.28*
	[0.84]	[0.24]	[0.99]	[2.00]	[1.74]	[1.05]	[2.15]
Corruption(*i*)	0.16	0.10	0.06	−0.01	0.29	−0.04	−0.32
	[0.32]	[0.60]	[0.22]	[0.07]	[1.52]	[0.24]	[1.59]
Corruption(*j*)	−0.32	0.59***	0.94***	−0.11	−0.19	−0.36*	−0.74***
	[0.65]	[3.91]	[3.98]	[0.77]	[1.17]	[2.35]	[3.90]
Corruption(*i*) * ln (GDPpercapita(*i*))	−0.02	−0.01	−0.02	0.00	−0.02	0.01	0.04^{+}
	[0.38]	[0.72]	[0.54]	[0.15]	[1.19]	[0.64]	[1.85]

(continued)

Table 3 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ISIC classification	Tobacco (314)	Textiles (321)	Footwear (324)	Industrial chemicals (351)	Rubber products (355)	Electric machinery (383)	Transport equipment (384)
Corruption(*j*) * ln (GDPpercapita(*i*))	0.03	−0.07***	−0.10***	0.01	0.02	0.04**	0.08***
	[0.65]	[4.22]	[3.85]	[0.73]	[1.22]	[2.58]	[3.96]
Constant	−21.23	−71.98***	−47.69***	−66.75***	−49.87***	−61.17***	−19.58**
	[1.42]	[12.56]	[5.22]	[13.60]	[8.50]	[11.19]	[2.98]
N	19494	77116	45755	77541	61405	79755	69395
R^2	0.777	0.870	0.828	0.874	0.858	0.891	0.840
adj. R^2	0.737	0.858	0.807	0.862	0.843	0.880	0.824
RMSE	1.52	1.17	1.31	1.18	1.15	1.25	1.51
Country-pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note $^+p < 0.1$, $^*p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; Absolute t statistics in brackets; Higher values of the corruption variable denote greater corruption. All regressions based on robust standard errors clustered within country pairs rubber products, electric machinery, transport equipment, and in the textiles industry. Higher income in the importing country is associated with greater imports in the rubber products, electric machinery, transport equipment industry, and reduced imports in the textiles industry. As in case of bilateral exports, imports in the tobacco industry can be explained by differences in factor endowment between countries.

The effect of corruption on bilateral imports differs significantly from its effects on exports in the different industries. We do not observe any statistically significant negative effect of domestic corruption on imports. On the other hand, in case of the transport equipment industry, there is some evidence that higher levels of domestic corruption is associated with higher imports in the high income countries and reduced imports in the low income countries. There is substantial heterogeneity in the effect of partner country corruption on imports. Greater corruption in the exporter country is found to reduce imports for the high income importers and increase imports for the low income importers in the textiles, and in the footwear industry. On the other hand, greater corruption in the exporter is found to be associated with higher imports among the high income importers and reduced imports for the low income importers in the electric machinery, and in the transport equipment industry.

Tables 4 and 5 report the marginal effect of corruption on bilateral exports and imports on countries with the highest and the lowest per capita incomes in our

Table 4 Marginal effect of corruption on bilateral exports

	Domestic corruption	Partner corruption
Tobacco (ISIC = 314)		
Low per capita GDP	0.19	0.11
	(0.17)	(0.21)
High per capita GDP	−0.18*	−0.17*
	(0.08)	(0.07)
Textiles (ISIC = 321)		
Low per capita GDP	0.19**	−0.10
	(0.06)	(0.07)
High per capita GDP	−0.14***	0.02
	(0.03)	(0.03)
Footwear (ISIC = 324)		
Low per capita GDP	0.08	−0.15
	(0.11)	(0.10)
High per capita GDP	0.03	−0.09*
	(0.05)	(0.04)
Industrial chemicals (ISIC = 351)		
Low per capita GDP	0.13*	0.18**
	(0.07)	(0.07)

(continued)

Table 4 (continued)

	Domestic corruption	Partner corruption
High per capita GDP	−0.12***	−0.004
	(0.03)	(0.03)
Rubber products (ISIC = 355)		
Low per capita GDP	0.43***	0.22*
	(0.08)	(0.09)
High per capita GDP	−0.22***	−0.002
	(0.04)	(0.03)
Electric machinery (ISIC = 383)		
Low per capita GDP	0.004	0.41***
	(0.08)	(0.08)
High per capita GDP	0.03	−0.08**
	(0.03)	(0.03)
Transport equipment (ISIC = 384)		
Low per capita GDP	0.09	0.29***
	(0.08)	(0.08)
High per capita GDP	−0.18***	−0.07*
	(0.04)	(0.03)

Note $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; Robust standard errors in parentheses; Low per capita GDP refers to (logarithm of) real GDP per capita = 6.19 and high per capita GDP refers to (logarithm of) real GDP per capita = 10.52

Table 5 Marginal effect of corruption on bilateral imports

	Domestic corruption	Partner corruption
Tobacco (ISIC = 314)		
Low per capita GDP	0.04	−0.11
	(0.17)	(0.18)
High per capita GDP	−0.05	0.04
	(0.09)	(0.08)
Textiles (ISIC = 321)		
Low per capita GDP	0.02	0.17**
	(0.06)	(0.05)
High per capita GDP	−0.04	−0.13***
	(0.03)	(0.03)
Footwear (ISIC = 324)		
Low per capita GDP	−0.04	0.34***
	(0.10)	(0.08)
High per capita GDP	−0.11*	-0.08^{+}
	(0.05)	(0.04)

(continued)

Table 5 (continued)

	Domestic corruption	Partner corruption
Industrial chemicals (ISIC = 351)		
Low per capita GDP	0.005	−0.04
	(0.06)	(0.05)
High per capita GDP	0.02	0.01
	(0.03)	(0.02)
Rubber products (ISIC = 355)		
Low per capita GDP	0.13*	−0.05
	(0.06)	(0.05)
High per capita GDP	0.03	0.04
	(0.04)	(0.03)
Electric machinery (ISIC = 383)		
Low per capita GDP	0.03	-0.09^{+}
	(0.06)	(0.05)
High per capita GDP	0.09*	0.09**
	(0.03)	(0.03)
Transport equipment (ISIC = 384)		
Low per capita GDP	−0.07	−0.24***
	(0.07)	(0.07)
High per capita GDP	0.11**	0.12**
	(0.04)	(0.04)

Note $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; Robust standard errors in parentheses; Low per capita GDP refers to (logarithm of) real GDP per capita = 6.19 and high per capita GDP refers to (logarithm of) real GDP per capita = 10.52

sample. The results are based on the fixed effects specifications reported in Tables 2 and 3. From Table 4, in the majority of the sectors, higher domestic and partner country corruption appears to reduce bilateral exports for the high income exporters. For example, one standard deviation increase in domestic corruption in the high income countries reduces exports from these countries by about 15 % (exp (0.138) − 1) * 100). On the other hand, similar increase in domestic corruption for the low income exporters increases exports from these countries by more than 20 % in the textiles industry. As discussed above, exports seem to be affected by corruption more than imports. We find that both domestic and partner country corruption have positive effects on imports to the high income countries in the electric machinery, and in the transport equipment industry, while there is negative impact of partner country corruption on the low income importers in these two industries.

4 Conclusions

Traditional theories of trade have focused on differences in technology and endowments, and on market structure to explain the patterns of trade. Recent literature has also highlighted the role of institutions in determining trade flows. This paper is in the latter vein. Corruption may affect different industries in different ways depending on how dependent these industries on contractual agreements, credit markets, or on government institutions like public sector banks or on licensing bodies. Allowing for this heterogeneity we find evidence of significant differences between sectors, and also between richer and poorer nations. Overall, exports seem to be more affected by corruption relative to imports, especially in richer nations. Corruption in the trading partners also seems to reduce exports. The heterogeneity in the effects of corruption on sector level trade flows calls for more work which can identify the channels through which corruption affects the pattern of comparative advantage.

Appendix 1

List of countries used in the fixed effect regressions:

Algeria, Argentina, Armenia, Australia, Austria, Azerbaijan, Bangladesh, Bulgaria, Bolivia, Brazil, Botswana, Canada, Chile, China, Cameroon, Colombia, Costa Rica, Cyprus, Czech Republic, Denmark, Ecuador, Egypt, El Salvador, Ethiopia, Finland, France, Gabon, Germany, Ghana, Greece, Guatemala, Hong Kong, Honduras, Hungary, Indonesia, India, Ireland, Iran, Iceland, Israel, Italy, Ivory Coast, Jordan, Japan, Kenya, Kuwait, Lithuania, Latvia, Morocco, Moldova, Mexico, Malta, Mongolia, Mozambique, Malawi, Malaysia, Nigeria, Netherlands, Norway, New Zealand, Oman, Pakistan, Panama, Peru, Philippines, Poland, Portugal, Romania, Russia, Senegal, Singapore, Slovakia, Slovenia, South Africa, South Korea, Spain, Sri Lanka, Sweden, Switzerland, Thailand, Trinidad and Tobago, Tunisia, Turkey, Tanzania, Uganda, Ukraine, United Kingdom, Uruguay, United States, Venezuela, Yemen.

References

Anderson JE (1979) A theoretical foundation for the gravity model. Am Econ Rev 69:106–116

Anderson JE, Marcouiller D (2002) Insecurity and the pattern of trade: an empirical investigation. The Review of Economics and Statistics 84:342–352

Anderson JE, van Wincoop E (2003) Gravity with gravitas: a solution to the border puzzle. Am Econ Rev 93:170–192

Belloc M (2006) Institutions and international trade: a reconsideration of comparative advantage. J Econ Surveys 20:3–26

Bergstrand J (1985) The gravity equation in international trade, some microeconomic foundations and empirical evidence. Rev Econ Stat 67:474–481

Bergstrand J (1989) The generalized gravity equation, monopolistic competition, and empirical evidence. Rev Econ Stat 71:143–153

Berkowitz D, Moenius J, Pistor K (2006) Trade, law, and product complexity. Rev Econ Stat 88:363–373

Deardorff A (1998) Determinants of bilateral trade: does gravity work in a neoclassical world? In: Frankel J (ed) The Regionalization of the World Economy. University of Chicago Press

de Groot H, Linders GM, Rietveld P, Subramanian U (2004) The institutional determinants of bilateral trade patterns. Kyklos 57:103–123

Feenstra RC, Markusen JR, Rose AK (2001) Using the gravity equation to differentiate among alternative theories of trade. Can J Econ 34:430–447

Francois J, Manchin M (2013) Institutions, infrastructure, and trade. World Dev 46:165–175

Head K, Mayer T, Ries J (2010) The erosion of colonial trade linkages after independence. J Int Econ 81:1–14

Helpman E, Krugman P (1985) Market structure and foreign trade. MIT Press, Cambridge, MA

Krugman P (1980) Scale economies, product differentiation, and the pattern of trade. Am Econ Rev 70:950–959

Levchenko AA (2007) Institutional quality and international trade. Rev Econ Stud 74:791–819

Marjit S, Mandal B, Roy S (2014) Trade openness, corruption and factor abundance: evidence from a dynamic panel. Rev Dev Econ 18:45–58

Markusen JR, Venables AJ (2000) The theory of endowment, intra-industry and multi-national trade. J Int Econ 52:209–234

Markusen JR (2013) Putting per-capita income back into trade theory. J Int Econ 90:255–265

Mohlmann LJ, Ederveen S, de Groot H, Linders GM (2010) Intangible barriers to trade: a sectoral approach. In: van Bergeijk PAG, Brakman S (eds) The gravity model in international trade: advances and applications. Cambridge University Press, Cambridge, pp 224–251

Mundlak Y (1978) On the pooling of time series and cross sectional data. Econometrica 56:69–86

Nicita A, Olarreaga M (2006) Trade, production and protection 1976–2004. World Bank Econ Rev 21:165–171

Nunn N (2007) Relationship-specificity, incomplete contracts, and the pattern of trade. Quart J Econ 122:569–600

Nunn N, Trefler D (2014) Domestic institutions as a source of comparative advantage. In: Gopinath G, Helpman E, Rogoff K (eds) Handbook of international economics, vol 4, North Holland, pp 263–315

Ranjan P, Lee JY (2007) Contract enforcement and international trade. Econ Politics 19:191–218

Rauch J (1999) Networks versus markets in international trade. J Int Econ 48:7–35

Trefler D (1995) The case of the missing trade and other mysteries. Am Econ Rev 85:1029–1046

van Bergeijk PAG, Brakman S (2010) The gravity model in international trade: advances and applications. Cambridge University Press, Cambridge

Enlargement Decisions of Regional Trading Blocs

Sunandan Ghosh

Abstract This paper tries to link the literature on equilibrium analyses of trading blocs with those analyzing the effect of policy centralization corresponding to widening and deepening choices faced by existing unions. In particular, this paper analyzes the decision of an existing bilateral free trade area regarding expansion vis-à-vis consolidation (transformation into customs union) or both simultaneously. The analytical framework employs a standard oligopolistic strategic trade model in a four-country world incorporating both technology and market size asymmetries. Formation of customs union between two similar initial members can be sustained as a subgame perfect Nash equilibrium in a technologically asymmetric world. However, when market sizes are asymmetric, there will exist multiple equilibria depending on the extent of largeness of non-member countries.

Keywords Expansion · Consolidation · Regional trading blocs

JEL Classifications C72 · F12 · F15

1 Introduction

Regional trading blocs (RTBs) do not remain static over time and a major topic of debate in the existing literature pertains to their enlargement decisions. Such analyses are relevant to understand the unprecedented proliferation of RTBs over the last two and a half decades and necessary for the welfare implications of bloc formations, particularly for developing nations.[1] Thus, the question of including

[1] A closer look into the agreements reported to the WTO till 15th June, 2014 will reveal that all the "emerging economies" as classified by the IMF (2012) are members of multiple trading arrangements with different degrees of economic cooperation (see Table 1 in the Appendix).

S. Ghosh (✉)
Centre for Development Studies, Thiruvananthapuram 695011, Kerala, India
e-mail: sunandan@cds.ac.in; sunandan@cds.edu

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_7

new member countries in an existing bloc versus deeper integration among the existing member countries has emerged time and again in the history of the development of dominant RTBs.[2] The evolution of the RTBs over time reflects the dynamic nature of their equilibrium structure, thus, making it an indispensable issue in the literature of regional economic integration. Hence, at the onset, it is important to highlight the various perspectives existing on this issue.

To start with, Kemp and Wan (1976) argues that an appropriate adjustment of external tariffs can create the possibility of constructing a path of Pareto-improving trading bloc formation, leading all the way to global free trade (GFT). Kennan and Riezman (1990) show that, by separating effects of tariff reduction and policy coordination, a movement from Nash equilibrium tariffs to a free trade area (FTA) improves global resource allocations. However, moving from FTA to customs union (CU) has ambiguous resource allocation effects.

Yi (2000) uses a linear-quadratic model of intra-industry trade to show that formation of FTAs induces member states to reduce external tariff levels sufficiently to make non-member countries better-off and hence, formation of FTA is a Pareto improvement in itself. However, as Yi argues, due to free-riding problems, GFT may not turn out to be a stable solution. Yi (1996) shows that the grand custom union is the unique pure-strategy Nash equilibrium outcome of an open regionalism game. However, it is typically not the equilibrium of a unanimous regionalism game.[3]

Aghion et al. (2007) on the other hand, use a dynamic bargaining model to evaluate the choice of sequential FTA formation vis-à-vis simultaneous multilateralism as faced by the leading negotiating potential member country. They argue that under grand coalition and in absence of externalities, GFT is the unique equilibrium. However, the authors also identify the conditions under which both sequential FTA formation and simultaneous multilateralism are equilibrium outcomes.

Park and Park (2008) use a combination of ex-ante simulation and ex-post econometric methods to analyze the choice of economic integration (FTA vs. CU) and finds that a CU with minimum common external tariff (CET) is the best option not only for the members but for the ROW as well. Seidmann (2009) uses a model of trade negotiation and a positioning game to answer the same question as Park and Park (2008) and finds that the level of "patience" of the concerned potential members plays the crucial role.

Freund (2000), one of the foremost works that considered expansion of an existing preferential trade agreement (PTA), claimed that regionalism provides

---

[2]Such a debate surfaced in the case of the EU at the time of providing accession to Mediterranean agricultural countries of Greece and Spain in the 1980s and, of late, to the East European countries. The evolution of the ASEAN reflects a similar story.

[3]Improving upon this result Yi (2000) shows that for homogeneous products, GFT is a pure-strategy Nash equilibrium outcome of the open regionalism game if and only if number of countries in the world is less than or equal to 9 (i.e., $N \leq 9$). Moreover, the global free trade area is the unique equilibrium outcome of the unanimous regionalism game for $2 < N < 8$.

first-mover advantage to the entrant nations, leading to higher welfare for the members of a PTA vis-à-vis non-members. However, Gil et al. (2008) do not find any such first-mover advantage using data for the European Union. The authors analyze the effect of both horizontal and vertical expansions using a gravity model. They conclude that enlargement of the EU has led to significant increase in trade volume while deepening has led to increased trade intensity among the EU member countries. They also find that integrating measures like formation of a common market and floating of the Euro have increased trade with the non-members countries. Lee et al. (2008) use panel data for 175 countries over the period 1948–99 and find that for a given regional trade arrangement (RTA), the non-members are always better-off by duplicating a separate RTA rather than joining the existing one. Egger and Larch (2008) use a panel for 145 countries over the period 1955–2005 and conclude that pre-existence of a PTA and geographical proximity increase the incentive for non-members to join an existing PTA or duplicate a separate PTA.

Although, the aforementioned studies discuss the stability of equilibrium structure, but, they don't exactly address the choice of an existing bloc over deepening or widening (alternatively, vertical and horizontal expansion) or both (simultaneous expansion). Hence, we need to focus on another set of literature—that analyzing the existence of any trade-off between the choices of widening vis-à-vis deepening in the context of multilateral institutions and international unions. The prime focus of this set of literature is policy centralization or policy-uniformity.

Gilligan (2004) shows that there exists a "broader-deeper trade-off" if the members opt for an "identical" policy; in absence of which there exists no such trade-off. Alesina et al. (2005) identify a similar trade-off stemming out of policy-uniformity. They argue that the trade-off between size and scope of international unions determine the optimal size of the unions—higher degree of policy centralization leads to smaller union size and vice versa. Hausken et al. (2006) consider a two-stage political economy game in line with unanimous regionalism to evaluate the conditions for widening vis-à-vis deepening. Similar to Alesina et al. (2005), they argue that expansion is preferred by the union members if lower degree of policy harmonization is required in the process of integration.

Lorz and Willmann (2008) use a political economy framework with strategic delegation. They observe that a critical level of lateral expansion of an economic union is required to dilute policy centralization. Before that threshold level, expansion is congenial to deeper cooperation.

Berglof et al. (2008) use a club-in-club theory and infer that the degree of integration is determined by the effort of the "weakest" or least efficient member under the assumption of unanimity. However, a threat of club-in-club can actually inspire the weaker members to exert more. They also argue that widening and deepening become compliments if new entrant has higher degree of efficiency.

This paper attempts to connect both these streams and shed some light into the choice of an existing RTB regarding inclusion of non-members vis-à-vis deeper integration among existing members. This has been analyzed in the light of an asymmetric world, both with respect to technology and domestic market sizes. Further, in case, the expansion or inclusion of a new member is preferred over consolidation of economic integration among the existing members, should an efficient or an inefficient non-member country be preferred? The above analysis has been executed using a theoretical framework of a four-country world-economy with a homogeneous good produced under monopolistic market conditions with linear inverse demand and cost functions. Possibility of trade in the identical good arises due to 'reciprocal dumping' motives of firms—strategic trade theory *a la* Brander (1981), Brander and Krugman (1983).[4] The point of analysis is how do such cross-country asymmetries influence the decision of an initial two-country RTB regarding further economic integration between them (the vertical expansion) vis-à-vis offering accession to a non-member country in the RTB (horizontal expansion) or both simultaneously. Given this framework, a CU formation between initial members will be sustained as the subgame perfect Nash equilibrium (SPNE) under technology asymmetry. However, there exists no unique SPNE given market size asymmetry. The equilibrium outcome depends on the market size of the bigger non-member country.

The paper is organized as follows. Section 2 describes the model. Section 3 analyzes the benchmark case under technology asymmetry. Section 4 finds the solution to the game of expansion. Section 5 tests the robustness of the model given market size asymmetry. Section 6 concludes the paper followed by appendices.

2 The Model

Consider a four-country world consisting of countries 1, 2, 3 and rest of the world (ROW). Each country has a single firm that produces a homogeneous good X locally and supplies in its own country-market (X_{ii}) as well as in other country-markets $[X_{ij}\,(i \neq j)]$.

Let us consider a linear inverse demand function[5] for the ith market

$$P_i = a_i - \sum_{j=1}^{4} X_{ji} \quad \forall \quad i, j = 1, 2, 3, 4. \tag{1}$$

where, X_{ji} is the supply of firm j in country-i market, P_i is the domestic price level in the ith country and the intercept term a_i gives us a measure of the market size.

[4]This theoretical framework revolves around the role of intra-industry trade (IIT). So far the emerging economies are concerned, IIT plays a major role in their trading patterns and such an observation is relevant for developing economies as well [see Brulhart (2009), Menon (1996), Khalifa (1996), Hurley (2003), Ahlstrom and Stalros (2005), Hu and Ma (1999) and Gaulier et al. (2012) for the importance of IIT in trade for emerging and developing nations across the globe].

[5]Thus, it is assumed that demand functions in different markets are different only with respect to maximum willingness-to-pay.

Markets are segmented by the assumption of constant marginal costs and cost functions are linear such that

$$C_i = c_i X_i = c_i \sum_{j=1}^{4} X_{ij} \quad \forall \quad i, j = 1, 2, 3, 4. \tag{2}$$

Hence, unlike Brander (1981), neither costs nor market sizes are identical.

Given the linear cost function defined in (2) and specific tariffs, the profit functions of the four national firms can be written as

$$\pi_i = \left(\sum_{j=1}^{4} P_j X_{ij} - c_i \sum_{j=1}^{4} X_{ij} - \sum_{k=1}^{4} t_k X_{ik} \right) \quad \forall \quad i, j, k = 1, 2, 3, 4; \quad i \neq k \tag{3}$$

The firms are assumed to be Cournot players so that they choose their output levels to maximize profits given the output choices of the rival firms.

Each national government protects the import competing firm by imposing specific tariff t_i on its imports $(X_{ji}; i \neq j)$ from the other three countries and chooses its tariff rates by maximizing national welfare given the tariff rates imposed by other governments.[6] Welfare function for the ith country is defined as the sum of the profit of the ith firm (π_i), the surplus accruing to the consumers of the ith country (CS_i) and the total tariff revenue generated by the ith country's imposition of specific tariffs on its imports (TR_i).

$$W_i = \pi_i + \mathrm{CS}_i + \mathrm{TR}_i \quad \forall \quad i = 1, 2, 3, 4 \tag{4}$$

Let us consider a two-stage game where the national governments choose optimum tariffs on each other's imports in the first stage and then in the second stage the firms choose their output levels (both domestic sales and exports). The relevant equilibrium concept is SPNE which is obtained through backward induction method.

3 The Benchmark Case: Technology Asymmetry

In the context of the above framework the alternative trade policy scenarios that are available to these four countries are as follows.

i. Non cooperatively chosen unilateral optimal tariffs where each of the countries imposes a specific tariff on its imports from others. The world system will be [1, 2, 3, 4].

---

[6]So, in a sense, the national governments also behave as Cournot players while deciding upon their optimal tariff levels.

ii. Two countries having an FTA among them while the other two operating unilaterally. The world trading system will be [(1 + 2), 3, 4].
iii. Expansion of the existing FTA via offering accession to any one of the other two countries, that is, horizontal expansion but not GFT. The world system will look like [(1 + 2 + 3), 4] or [(1 + 2 + 4), 3].
iv. Customs union (CU) formation among the initial member states (vertical expansion), that is, [{1 + 2}, 3, 4].
v. CU formation among the members of the enlarged FTA (simultaneous expansion), that is, [{1 + 2 + 3}, 4] or [{1 + 2 + 4}, 3].

Let's begin with the benchmark case with countries 1 and 2 having a FTA among them with Country-3 and the ROW operating in isolation [(1 + 2), 3, 4]. Hence, countries 1 and 2 do not impose any tariff on imports from each other but continue to unilaterally impose tariffs on their respective imports from Country-3 and the ROW.

At this point let us assume that market sizes of all the four countries are identical and the national firms of countries 1 and 2 have equally efficient production technology as compared to a backward production technology in the rest. In particular, we assume $a_i = a \, \forall \, i = 1, 2, 3, 4$ and $c_1 = c_2 = c < c_3 \neq c_4$.

Starting with this benchmark model can be justified in light of the following analytical reasons. First, introducing differences in market size later will then clearly bring out the implications for expansions of an existing FTA. Second, any further expansion of the initial FTA (horizontal or vertical or both) will be different from GFT. Third, if accession is offered by the existing members to any one of the non-members, then to whom it will be offered (efficient or inefficient) is clearly distinguishable from the above assumption of $c_3 \neq c_4$. Finally, since, primary focus of this paper is not how and why a FTA is being formed [as in Raff (2001)], the above assumption enables us to assume that an initial two-country FTA was formed between the countries with similar technological advancement.

Profit functions as given in (3) change only in respect of t_1 and t_2 in π_1 and π_2 as countries 1 and 2 do not impose any tariff on imports from each other. Profit maximization yields the domestic supply (X_{ii}) and import levels (X_{ji}) in each country as follows

$$X_{ij} = \frac{1}{5}\left(a - 3c + c_3 + c_4 + 2t_j\right) \quad \forall \quad i = 1, 2; j = 1, 2 \tag{5a}$$

$$X_{ij} = \frac{1}{5}\left(a + 2c - 4c_3 + c_4 - 3t_j\right) \quad \forall \quad i = 3; j = 1, 2 \tag{5b}$$

$$X_{ij} = \frac{1}{5}\left(a + 2c + c_3 - 4c_4 - 3t_j\right) \quad \forall \quad i = 4; j = 1, 2 \tag{5c}$$

$$X_{ij} = \frac{1}{5}\left(a - 3c + c_3 + c_4 - 2t_j\right) \quad \forall \quad i = 1, 2; j = 3, 4 \tag{5d}$$

$$X_{ij} = \frac{1}{5}\left(a + 2c - 4c_i + c_j - 2t_j\right) \quad \forall \quad i = 3,4; j = 3,4; \quad i \neq j \tag{5e}$$

$$X_{33} = \frac{1}{5}(a + 2c - 4c_3 + c_4 + 3t_3) \tag{5f}$$

$$X_{44} = \frac{1}{5}(a + 2c + c_3 - 4c_4 + 3t_4) \tag{5g}$$

Putting the above values of the X_{ij}'s in (1) we have the market clearing prices as

$$P_i = \frac{1}{5}(a + 2c + c_3 + c_4 + 2t_i) \quad \forall \quad i = 1,2 \tag{6a}$$

$$P_i = \frac{1}{5}(a + 2c + c_3 + c_4 + 3t_i) \quad \forall \quad i = 3,4 \tag{6b}$$

Now, the ith country takes t_{-i} to be given and maximizes the national welfare W_i given by (4) with respect to t_i. Given the FTA$_{12}$ regime, such welfare maximization yields the following optimal tariff levels.[7]

$$\tilde{t}_i^{12} = \frac{1}{16}(2a + 4c - 3c_3 - 3c_4) \quad \forall \quad i = 1,2 \tag{7a}$$

$$\tilde{t}_i^{12} = \frac{1}{33}(9a - 2c - 6c_i - c_j) \quad \forall \quad i = 3,4; j = 3,4; \quad i \neq j \tag{7b}$$

Note that, from now onwards the optimal tariff levels and equilibrium quantities, prices and welfare levels will be denoted in the following manner—tilde ($\sim$) for an FTA regime and hat (^) for a CU regime, superscript will denote the members of the trade bloc and subscripts will denote the country (countries) concerned.

4 Expansion of the Existing FTA

This section analyzes the choice of the member countries of the existing FTA to expand or to consolidate their cooperation. The timing of decision making process is illustrated by the decision tree in Fig. 1. One option is to include Country-3 (or the ROW) as a new member to the existing FTA. The other option is to consolidate the cooperation between the existing members (Country-1 and Country-2) by forming a customs union, whereby the countries unify their tariff policy and cooperatively impose a CET on the non-members. Another option for the existing members is to expand (by giving accession to a new member) and consolidate

[7] See Appendix 1.

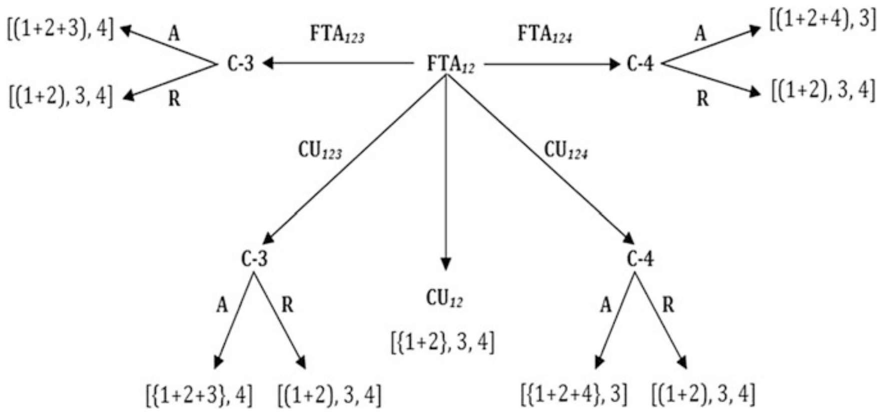


Fig. 1 The game tree

simultaneously. However, for both lateral and simultaneous expansions, it is necessary for Country-3 (or the ROW) to find it optimal to enter the existing FTA. The solution is obtained following backward induction.[8]

4.1 *Second-Stage Decisions*

As evident from the game tree, we have to distinguish between two alternative sets of second-stage decisions for non-member countries. First, when offered accession to FTA_{12}, they will have to decide whether to accept or reject. Second, if instead FTA_{12} consolidates into a CU_{12} and offers accession to it, again the decision is whether to accept it or reject. To begin with, let us analyze the case when FTA_{12} plans a lateral expansion by offering accession to the non-members (Country-3 and the ROW).

4.1.1 Horizontal Expansion and the Non-member Country's Decision

When Country-3 is offered accession to FTA_{12} by the existing member countries and it accepts the offer, a new FTA_{123} is formed with only the ROW being left out.

[8]We start with the second-stage decisions of the non-member countries when offered accession to FTA_{12}—resulting in FTA_{123} or FTA_{124} as the case may be—and when offered accession to CU_{12} being formed by countries 1 and 2—resulting in a larger customs union, CU_{123} or CU_{124} as the case may be. Given such choices, we fold back to the first-stage decision of the member countries over FTA_{123}, FTA_{124}, CU_{12}, CU_{123}, CU_{124} or remaining as FTA_{12}.

Profit maximization[9] will yield optimal domestic supply and import levels in each country which in turn will generate the market clearing prices.[10] Now, welfare maximization yields the following optimal tariff levels

$$\tilde{t}_i^{123} = \frac{1}{37}(3a + 6c + 8c_3 - 17c_4) \quad \forall \quad i = 1, 2 \tag{8a}$$

$$\tilde{t}_3^{123} = \frac{1}{37}(3a + 16c - 2c_3 - 17c_4) < \tilde{t}_3^{12} \tag{8b}$$

$$\tilde{t}_4^{123} = \frac{1}{33}(9a - 2c - c_3 - 6c_4) = \tilde{t}_4^{12} \tag{8c}$$

Before proceeding further, two comments should be made at this point. First, the non-member ROW continues to impose the same level of unilaterally optimal tariff on its imports from the other three countries as compared to the unilaterally optimal tariff it used to impose under FTA_{12} regime. Second, as Country-3 accepts the offer of accession extended to it by the existing member countries 1 and 2, Country-3 continues to impose unilaterally optimal tariff on its imports from the ROW only. However, such tariff rate, as evident from (8b), is lower than the level of tariff it used to impose on its imports under FTA_{12} regime. Hence, formation of RTB (FTA_{123} in this case) reduces the unilaterally optimal tariff for the new entrant country (Country-3 in this case). This is because of the fact that tariff protection of an inefficient domestic producer comes with some welfare costs. Hence, for the new entrant, lowering tariff on its imports from the non-member will reduce the consumption distortions and hence, in turn, increase the welfare.

Now, if given an offer of accession, Country-3 will decide upon accepting or rejecting the offer depending upon the levels of national welfare it faces under the alternative regimes, that is, it will compare $\tilde{W}_3^{123}$ and $\tilde{W}_3^{12}$ in particular. Relevant calculations[11] reveal that

$$\begin{aligned} \tilde{W}_3^{123} - \tilde{W}_3^{12} &= \tilde{\Delta}_3 \\ &= 0.018a^2 + 0.323ac - 0.252ac_3 - 0.107ac_4 + 0.385c^2 - 0.671cc_3 \\ &\quad - 0.422cc_4 + 0.18c_3^2 + 0.562c_3c_4 - 0.016c_4^2 \end{aligned} \tag{9}$$

---

[9]Under FTA_{123} regime none of Countries 1, 2, and 3 imposes any tariff on imports from each other. Hence, profit functions for Countries 1, 2, and 3 will change accordingly. However, the ROW continues to impose tariff on its imports from countries 1, 2, and 3 and its profit function will remain unaltered.

[10]The exercise is exactly similar to that done while obtaining (5a)–(6b).

[11]See Appendix 2.

It is evident from (9) that the sign of $\tilde{\Delta}_3$ is ambiguous and hence, whether the national welfare of Country-3 is higher under FTA$_{123}$ regime or not depends, among others, on the magnitude of technology asymmetries of countries. To fix the idea, suppose $c_3 = c_4$, that is, firms belonging to the non-member countries are equally efficient. In such a case, (9) boils down to

$$\tilde{\Delta}_3 = 0.018a^2 + 0.323ac - 0.359ac_3 + 0.385c^2 - 1.093cc_3 + 0.726c_3^2 \qquad (10)$$

Thus, the decision of Country-3 in joining FTA$_{12}$ depends on its firm's cost inefficiency relative to the firms in countries 1 and 2. To trace out how the welfare change varies with such relative cost inefficiency, we fix the value of c at an arbitrary level and raise $c_3 = c_4$ from such a level. As shown in Appendix 3, such a relationship is non-monotonic over the relevant range $c_3 \in [c, a]$. Figure 2 illustrates this non-monotonic relationship. Note that if $c_3 = c_4 > a$, then the firms in Country-3 and the ROW cannot survive even when monopoly power is protected in their respective domestic markets. On the other hand, by definition $c_3 = c_4 > c$. Thus, Country-3 will accept the offer of accession to the existing FTA$_{12}$ when its firm is very similar such that $c_3 \in (c, \tilde{\tilde{c}})$.

The economic intuition behind such decision can be explained with the help of the welfare function defined in (4). As Country-3 joins the FTA, it forgoes a part of the tariff revenue of the magnitude $(X_{13} + X_{23})\tilde{t}_3^{12}$ as it does not impose tariff on its imports from countries 1 and 2 under FTA$_{123}$ regime. This is the loss from joining the existing FTA$_{12}$. Also pro-competitive effect of joining FTA$_{12}$ causes profit of the firm in Country-3 to fall in its domestic market. These two losses have to be weighed against two sources of gains from joining FTA$_{12}$. First, as countries 1 and 2 do not impose any tariff on their imports from Country-3 under the FTA$_{123}$ regime, the profit of firm-3 from exporting the good in FTA$_{12}$ market increases. Second, consumers in Country-3 gain because Country-3 now removes tariff on imports from FTA$_{12}$. When firm-3 is less inefficient in the sense that $c < c_3 < \tilde{\tilde{c}}$, its

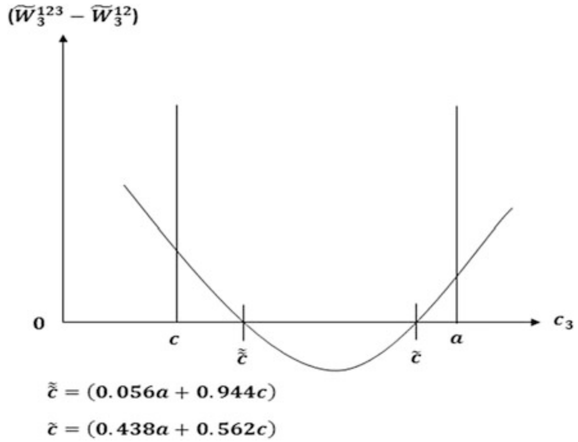


Fig. 2 Decision of country-3 (FTA$_{123}$)

profit gain in Country-1 and 2 markets are much larger than the loss of profit in its own market. This, along with larger consumer surplus, over-compensates the loss of tariff revenue for Country-3. For moderately high cost inefficiency profit loss from domestic sales increases whereas profit gain from exports falls. This together with tariff revenue loss makes the option of joining FTA$_{12}$ welfare reducing. It should be noted that $\tilde{\Delta}_3 > 0$ when $c_3 > \tilde{c}$. But, the exports of Country-3 to the initial members are positive as long as $c_3 < \left(\frac{a+2c}{3}\right)$. Hence, the zone $c_3 \in \left(\frac{a+2c}{3}, \tilde{c}\right)$ becomes irrelevant.[12]

Now we consider, the other branch of the game tree where the ROW is being offered accession by countries 1 and 2 for an enlarged FTA, that is, the world system will look like [(1 + 2 + 3), 4]. Replicating the method of calculation we arrive at the welfare maximizing optimal tariff levels

$$\tilde{t}_i^{124} = \frac{1}{37}(3a + 6c - 17c_3 + 8c_4) \quad \forall \quad i = 1,2 \tag{11a}$$

$$\tilde{t}_4^{124} = \frac{1}{37}(3a + 16c - 17c_3 - 2c_4) < \tilde{t}_4^{12} \tag{11b}$$

$$\tilde{t}_3^{124} = \frac{1}{33}(9a - 2c - 6c_3 - c_4) = \tilde{t}_3^{12} \tag{11c}$$

Comparing the set of tariffs as obtained in (10) and (11) it is evident that

$$\tilde{t}_i^{123} - \tilde{t}_i^{124} = \frac{25}{37}(c_3 - c_4) \quad \forall \quad i = 1,2 \tag{12a}$$

$$\Rightarrow \quad \left(\tilde{t}_i^{123} - \tilde{t}_i^{124}\right) \lesseqqgtr 0 \text{ if } \left(c_3 \lesseqqgtr c_4\right) \tag{12b}$$

$$\tilde{t}_3^{123} - \tilde{t}_4^{124} = \frac{15}{37}(c_3 - c_4) \tag{12c}$$

$$\Rightarrow \quad \left(\tilde{t}_3^{123} - \tilde{t}_4^{124}\right) \lesseqqgtr 0 \text{ if } \left(c_3 \lesseqqgtr c_4\right) \tag{12d}$$

$$\tilde{t}_4^{123} - \tilde{t}_3^{124} = \frac{5}{33}(c_3 - c_4) \tag{12e}$$

$$\Rightarrow \quad \left(\tilde{t}_4^{123} - \tilde{t}_3^{124}\right) \lesseqqgtr 0 \text{ if } \left(c_3 \lesseqqgtr c_4\right) \tag{12f}$$

Before proceeding with this analysis further, some observations regarding the optimal tariff levels are warranted. First, comparing (8b) and (11b) it is easily evident that the new entrant country reduces its tariff on imports from the non-member country. Second, the non-member country continues to impose the

[12]Note that $\tilde{\tilde{c}} < \frac{a+2c}{3} < \tilde{c}$.

same unilaterally optimal tariff level on its imports. Third, from (12d) and (12f) it is evident that more efficient the country is in terms of production technology, the lower will be its unilaterally optimal tariff.[13]

Now, if given offer of accession, the ROW will decide upon accepting or rejecting the offer depending upon the levels of national welfare it faces under the alternative regimes, that is, it will compare $\tilde{W}_4^{124}$ and $\tilde{W}_4^{12}$. Calculations similar to that in Appendix 2 reveal that

$$\tilde{W}_4^{124} - \tilde{W}_4^{12} = 0.018a^2 + 0.323ac - 0.107ac_3 - 0.252ac_4 + 0.385c^2 - 0.422cc_3 - 0.671cc_4 - 0.016c_3^2 + 0.562c_3c_4 + 0.18c_4^2 \tag{13}$$

It is easily evident that Eqs. (9) and (13) are absolutely similar[14] and hence, we can claim that the ROW, just like Country-3, will accept the offer of accession to the existing FTA$_{12}$ as long as it is almost as efficient as the initial FTA members.

Lemma 1 *Both the non-members will accept offer of accession extended by the existing members to form an enlarged FTA when their production technology is very similar to that of the existing members in the sense that* $c_i \in (c, \tilde{\tilde{c}}) \quad \forall \quad i = 3, 4$.

4.1.2 Simultaneous Consolidation and Expansion and the Non-member Country's Decision

In this section we consider the optimal decision of the non-member countries when countries 1 and 2 consolidate into a CU[15] and at the same time offers accession to any one of the non-member countries. Thus, if the non-member country getting the offer of accession accepts it, we have either CU$_{123}$ or CU$_{124}$ as the case may be.

To start with, we consider the case of Country-3 being offered accession by the existing member countries to form an enlarged CU with only the ROW being left out. Under CU$_{123}$ regime Countries 1, 2, and 3 impose zero tariffs on imports from each other but a CET on their imports from the ROW. Given this regime, the profit functions will look like

[13]This is because of the fact that tariff protection of an inefficient domestic producer comes with some welfare costs. That is, the welfare cost is larger if an efficient importer is prohibited and hence, the welfare maximizing optimal tariff level is lower as the national firm is more efficient.

[14]The results are similar due to the similarity in the underlying structure given by the profit functions and optimal tariff levels. Moreover, the decision of the non-member ROW has been calculated in a manner exactly similar to that done for Country-3 as shown in Appendix 3.

[15]Remember that the difference between CU and FTA is the CET to be set by the CU bloc members (joint welfare maximization) vis-à-vis unilaterally optimal tariff set by the FTA bloc members on imports from non-members.

$$\pi_i = \left(\sum_{j=1}^{4} P_j X_{ij} - c_i \sum_{j=1}^{4} X_{ij} - t_4 X_{i4} \right) \quad \forall \quad i = 1, 2, 3; j = 1, 2, 3, 4 \tag{14a}$$

$$\pi_4 = \left(\sum_{j=1}^{4} P_j X_{4j} - c_4 \sum_{j=1}^{4} X_{4j} - \hat{t} \sum_{i=1}^{3} X_{4i} \right) \quad \forall \quad i = 1, 2, 3; j = 1, 2, 3, 4 \tag{14b}$$

where, $\hat{t}$ is the CET imposed by members of CU_{123} on their imports from the ROW. Welfare maximizing tariff[16] levels turn out to be

$$\hat{t}_{123} = \frac{1}{33}(7a + 4c + 2c_3 - 13c_4) \tag{15a}$$

$$\hat{t}_4^{123} = \frac{1}{33}(9a - 2c - c_3 - 6c_4) = \tilde{t}_4^{123} \tag{15b}$$

Now, if given offer of accession, Country-3 will decide upon accepting or rejecting the offer depending upon the levels of national welfare it faces under the alternative regimes of CU_{123} and FTA_{12}. Comparing $\hat{W}_3^{123}$ and $\tilde{W}_3^{12}$ we have (see Appendix 5)

$$\begin{aligned} \hat{W}_3^{123} - \tilde{W}_3^{12} = \hat{\Delta}_3 = & 0.03a^2 + 0.421ac - 0.383ac_3 - 0.096ac_4 + 0.3c^2 - 0.645cc_3 \\ & - 0.373cc_4 + 0.267c_3^2 + 0.496c_3c_4 - 0.013c_4^2 \end{aligned} \tag{16}$$

Once again the sign of $\hat{\Delta}_3$ is ambiguous and we proceed by assuming some arbitrary values for c_3 and c_4. To start with, we assume $c_3 = c_4$ and (16) boils down to

$$\hat{\Delta}_3 = 0.03a^2 + 0.421ac - 0.479ac_3 + 0.3c^2 - 1.018cc_3 + 0.75c_3^2 \tag{17}$$

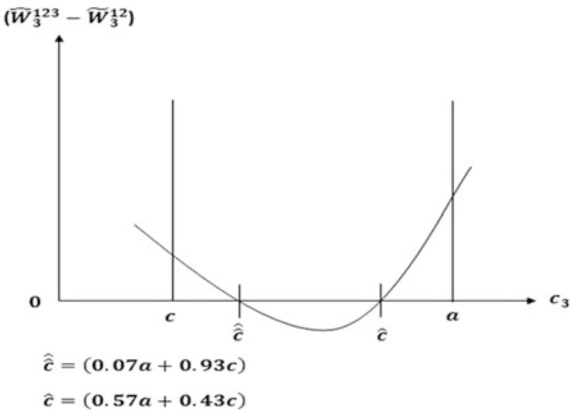
As shown in Appendix 6, such a relationship is non-monotonic over the relevant range $c_3 \in [c, a]$.[17] Figure 3 illustrates this non-monotonic relationship.

Once again, as in the case of lateral expansion of the FTA, Country-3 will accept the offer of accession to the existing FTA_{12} when its firm is very similar such that $c_3 \in \left(c, \hat{\hat{c}}\right)$. The economic intuition behind such decision is analogous to that described earlier. As Country-3 joins the CU, it forgoes a part of the tariff revenue of the magnitude $(X_{13} + X_{23})\tilde{t}_3^{12}$ as it does not impose tariff on its imports from countries 1 and 2 under CU_{123} regime. Moreover, comparing (7b) and (15a) we

[16]See Appendix 4 for the derivation of $\hat{t}_{123}$ using joint welfare maximization.

[17]Exports of Country-3 ceases as $c_3 > \frac{a+2c}{3}$ and $\hat{\hat{c}} < \frac{a+2c}{3} < \hat{c}$.

Fig. 3 Decision of country-3 (CU_{123})



have $\left(\tilde{t}_3^{12} - \tilde{t}_{123}\right) = \frac{2}{33}(a - 3c - 4c_3 + 6c_4) > 0$ for $c_3 = c_4$. Hence, the tariff revenue accruing to Country-3 on the imports from the ROW declines as well. Therefore, when Country-3 is less inefficient $\left\{c_3 \in \left(c, \hat{\hat{c}}\right)\right\}$, its profit gain is larger than the loss of profit in its own market. This along with larger consumer surplus over-compensates the loss of tariff revenue for Country-3.

Next, we consider the other case where the ROW is being offered accession an enlarged CU, that is, a world that looks like [{1 + 2 + 4}, 3]. The optimal tariff levels are calculated to be

$$\hat{t}_{124} = \frac{1}{33}(7a + 4c - 13c_3 + 2c_4) \tag{18a}$$

$$\hat{t}_3^{124} = \frac{1}{33}(9a - 2c - 6c_3 - c_4) = \hat{t}_3^{124} \tag{18b}$$

Now, if given offer of accession, the ROW will decide upon accepting or rejecting the offer depending upon the levels of national welfare it faces under the alternative regimes of CU_{124} and FTA_{12}. Calculations, exactly similar to that done in Appendix 6 for Country-3, reveal that

$$\begin{aligned} \hat{W}_4^{124} - \tilde{W}_4^{12} = {} & 0.03a^2 + 0.421ac - 0.096ac_3 - 0.383ac_4 + 0.3c^2 - 0.373cc_3 \\ & - 0.645cc_4 - 0.013c_3^2 + 0.496c_3c_4 + 0.267c_4^2 \end{aligned} \tag{19}$$

Equations (16) and (19) are again similar and, as argued in the previous subsection, we get exactly same results for ROW as we had obtained for Country-3. Thus, in this case also, we can claim that the ROW, just like Country-3, will accept the offer of accession to form an enlarged CU as long as its firm is very similar such that $c_4 \in (c, \hat{\hat{c}})$.

Lemma 2 *Both the non-members will accept offer of accession extended by the existing members to form an enlarged CU with the existing FTA_{12}members given their production technology is very similar to that of the existing members in the sense that* $c_i \in (c, \hat{\hat{c}}) \quad \forall \quad i = 3, 4$.

4.2 First-Stage: Optimal Expansion Decision of Member Countries

Given the optimal choices of the non-member countries when they are offered accession to FTA_{12} as specified in Lemmas 1 and 2, we now examine the optimal decisions of the member countries of FTA_{12}. This decision is important because even if the non-members decide unilaterally to join the existing FTA_{12}, whether any such expansion will occur or not depends solely on the decisions of the existing members. On the one hand, given open regionalism, any expansion of the existing FTA will occur if the non-members decide to join. On the other hand, given unanimous regionalism, any further expansion of the existing FTA will occur only if the members offer accession to any non-member country. However, in reality, open regionalism is hardly followed. Therefore, while determining the first-stage decisions, this paper assumes unanimous regionalism.

The member countries will compare their respective national welfare levels across all the possible expansion options (shown in Fig. 1) and choose that expansion strategy which yields highest welfare level. Note that, we will explain the decision of the member countries with the case of only Country-1 throughout this section as the decisions for both Country-1 and Country-2 will exactly be the same.[18]

We start with the benchmark case of FTA_{12} as discussed in Sect. 2. Using (5a)–(7b) we have the welfare level of Country-1 under the benchmark case of FTA_{12} to be

$$\begin{aligned}\tilde{W}_1^{12} = 0.454a^2 - 0.959ac + 0.025ac_3 + 0.025ac_4 + 1.413c^2 - 0.933cc_3 \\ - 0.933cc_4 + 0.228c_3^2 + 0.45c_3c_4 + 0.228c_4^2\end{aligned} \tag{20}$$

The welfare levels of Country-1 under the lateral expansion possibilities[19] turn out to be

[18]Given the assumption that countries 1 and 2 are exactly similar with respect to both market size and production technology, the welfare levels for both of them will be exactly the same.

[19]See Appendix 7.

$$\tilde{W}_1^{123} = 0.467a^2 - 1.113ac + 0.171ac_3 + 0.008ac_4 + 1.367c^2 - 0.823cc_3 \\ - 0.818cc_4 + 0.234c_3^2 + 0.182c_3c_4 + 0.314c_4^2 \tag{21}$$

$$\tilde{W}_1^{124} = 0.467a^2 - 1.113ac + 0.008ac_3 + 0.171ac_4 + 1.367c^2 - 0.818cc_3 \\ - 0.823cc_4 + 0.314c_3^2 + 0.182c_3c_4 + 0.234c_4^2 \tag{22}$$

Similarly, the welfare levels of Country-1 under the simultaneous consolidation and expansion possibilities (that is CU$_{123}$ and CU$_{124}$) turn out to be[20]

$$\hat{W}_1^{123} = 0.478a^2 - 1.196ac + 0.22ac_3 + 0.019ac_4 + 1.453c^2 - 0.85cc_3 \\ - 0.859cc_4 + 0.217c_3^2 + 0.207c_3c_4 + 0.317c_4^2 \tag{23}$$

$$\hat{W}_1^{124} = 0.478a^2 - 1.196ac + 0.019ac_3 + 0.22ac_4 + 1.453c^2 - 0.859cc_3 \\ - 0.85cc_4 + 0.317c_3^2 + 0.207c_3c_4 + 0.217c_4^2 \tag{24}$$

The only remaining possibility needed to be analyzed is the option of consolidation without any lateral expansion. When Countries 1 and 2 form a customs union only between them, they together impose a CET (obtained via joint welfare maximization) on their imports from Country-3 and the ROW. Proceeding as before, welfare maximization yields the following optimal tariff levels[21]

$$\hat{t}_i^{12} = \frac{1}{8}(2a - c_3 - c_4) > \hat{t}_i^{12} \quad \forall \quad i = 1, 2 \tag{25}$$

$$\hat{t}_i^{12} = \frac{1}{33}\left(9a - 2c - 6c_i - c_j\right) = \hat{t}_i^{12} \quad \forall \quad i = 3, 4; j = 3, 4; \quad i \neq j \tag{26}$$

The welfare level of Country-1 is given by[22]

$$\hat{W}_1^{12} = 0.598a^2 - 0.866ac + 0.214ac_3 + 0.214ac_4 + 0.807c^2 - 0.459cc_3 \\ - 0.459cc_4 + 0.232c_3^2 + 0.456c_3c_4 + 0.232c_4^2 \tag{27}$$

Now we compare the welfare levels for Country-1 under CU$_{12}$ regime with those under alternative regimes as defined and analyzed in the previous sections. For the purpose we compare (27) with (20), (21), (22), (23), and (24) separately and, as shown in Appendix 9, the welfare of Country-1 is highest under CU$_{12}$ regime. Same will be the case for Country-2 given the assumption that initial member countries are symmetric in both market size and production technology.

---

[20]See Appendix 8.

[21]See Appendices 1 and 4.

[22]Note that, given $a_1 = a_2$ and $c_1 = c_2$, we have $\hat{W}_1^{12} = \hat{W}_2^{12}$.

Lemma 3 *The welfares of both the initial* FTA_{12}*member countries are highest when they form a customs union among them.*

4.3 Subgame Perfect Expansion of the FTA

Now we are in a position to establish the SPNE of the game. From Lemmas 1 and 2 we know that both the non-member countries will accept any offer of accession from the existing members to form either an enlarged FTA or CU given certain parametric restrictions. However, the SPNE of the game depends on the fact that given unanimous regionalism, the existing members will decide on whether to offer accession to any new member as compared to consolidation among themselves. Now, given Lemma 3 we know that the existing member countries experience highest welfare when they go for deeper economic cooperation among themselves without giving accession to any new member. Hence, the existing member countries will opt for a customs union among them CU_{12} will be sustained as the SPNE of the game illustrated in the game tree (Fig. 1).

Proposition 1 *When there exists an FTA between two similar countries and there exist two inefficient non-member countries, then a customs union among the initial members will be the subgame perfect equilibrium outcome when the FTA members decide to expand.*

Proof Follows from Appendix 9.

Hence, the optimal equilibrium structure of a regional economic integration is a customs union among two similar countries. Such a result is partially in line with the results of Park and Park (2008) where the authors find it optimal for the member countries to implement a CU vis-à-vis an FTA. Yi (1996) argues that a grand customs union can be sustained as the unique pure-strategy Nash equilibrium given open regionalism. Yi (1996), however, does not suggest any unambiguous solution for unanimous regionalism game. Our result suggests that a customs union among two similar countries will be sustained as the SPNE under unanimous regionalism when there is no market or demand asymmetry whatsoever.

Given Lemma 3, deeper integration among existing members yields higher welfare, and hence, they opt for CU formation. Formation of CU requires higher degree of policy coordination among the member countries. Therefore, as the existing members opt for deeper integration and hence, higher degree of policy coordination, they prefer not to offer accession to new entrants. Such a result is also in line with Gilligan (2004), Alesina et al. (2005), Hausken et al. (2006).

5 Market Size Asymmetry

This section reconsiders the issue of expansion of the existing FTA between symmetric member countries under market size asymmetry. Like Sect. 2 we begin with the benchmark case with countries 1 and 2 having an FTA among them with Country-3 and the ROW being left to operate in isolation. At this point we make the assumption that all the four countries have identical production technology. However, the market sizes of only the existing members are identical. In particular, we assume $c_i = c \, \forall \, i = 1, 2, 3, 4$ and $a_1 = a_2 = a \neq a_3 > a_4$. The additional assumption of Country-3 having a larger market size helps us to unambiguously rank the welfare levels later. Given, the new set of assumptions, welfare maximizing optimal tariff levels turn out to be[23]

$$\tilde{t}_i^{12} = \frac{1}{8}(a - c) \quad \forall \quad i = 1, 2 \tag{28a}$$

$$\tilde{t}_i^{12} = \frac{3}{11}(a_i - c) \quad \forall \quad i = 3, 4 \tag{28b}$$

5.1 Second-Stage Decisions

The two alternative sets of second-stage decisions taken by the non-members (choice between accepting and rejecting the offer of accession to the existing FTA to form an enlarged FTA or CU) are analyzed in this section.

5.1.1 Horizontal Expansion and the Non-member Country's Decision

We start with the case when Country-3 is offered accession to FTA$_{12}$ by the existing member countries and it accepts the offer. The welfare maximizing optimal tariff levels are calculated in the same manner as in the earlier sections.

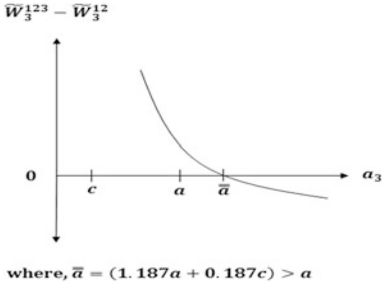
$$\tilde{t}_i^{123} = \frac{3}{37}(a - c) < \tilde{t}_i^{12} \quad \forall \quad i = 1, 2 \tag{29a}$$

$$\tilde{t}_3^{123} = \frac{3}{37}(a_3 - c) < \tilde{t}_3^{12} \tag{29b}$$

$$\tilde{t}_4^{123} = \frac{3}{11}(a_4 - c) = \tilde{t}_4^{12} \tag{29c}$$

[23] See Appendix 10.

Fig. 4 Decision of country-3



Now, as discussed in Sect. 4.1.1, here also we find that the new entrant country reduces its tariff on imports from the non-member country and the non-member country continues to impose the same unilaterally optimal tariff level on its imports.

Now, if given an offer of accession, Country-3 will decide upon accepting or rejecting the same by comparing $\tilde{W}_3^{123}$ with $\tilde{W}_3^{12}$. Relevant calculations (shown in Appendix 11) reveal that

$$\tilde{W}_3^{123} - \tilde{W}_3^{12} = \tilde{\Delta}_3 = 0.062a^2 - 0.044a_3^2 + 0.018c^2 + 0.088a_3c - 0.124ac \quad (30)$$

It is evident from (30) that the sign of $\tilde{\Delta}_3$ is ambiguous and hence, whether the national welfare of Country-3 is higher under FTA$_{123}$ regime or not depends, among others, on the magnitude of market size asymmetries of countries. As shown in Appendix 12, the relationship as given by (30) is monotonically decreasing[24] and Country-3 will accept the offer of accession as long as its market size is less than the critical value $\bar{\bar{a}} = (1.187a + 0.187c) > a$. This is illustrated in Fig. 4.

Now, such monotonic decline in aggregate welfare of Country-3 as it joins FTA$_{12}$ and forms an enlarged FTA$_{123}$ can be explained as follows. As Country-3 joins the FTA, it foregoes a part of its tariff revenue. Further, the profit of its firm reduces as a part of its sales fall in the domestic market due to competition from equally efficient firms of countries 1 and 2. Now, as it imposes no tariff on its imports from countries 1 and 2, consumer surplus increases. However, the larger the market size of Country-3 is, the larger will be the decline in the profit of its national firm in the domestic market. Hence, when its market size is considerably larger than that of countries 1 and 2, in the sense that $a_3 > \bar{\bar{a}} > a$, the increase welfare of Country-3 via increase in consumer surplus will fail to compensate the loss accruing to its national firm in the domestic market and, as a result, Country-3 will reject the offer of accession to the existing FTA$_{12}$.

Now, we consider the other branch of the game tree (as illustrated in Fig. 1) when the ROW is offered accession to FTA$_{12}$ by the existing member countries

[24]It may be noted that this monotonic relationship between market size asymmetry and welfare gain of Country-3 from joining the expanding FTA is a major difference with the cost asymmetry case discussed earlier.

with only Country-3 being left out. Similarly, the ROW will compare $\tilde{W}_4^{124}$ and $\tilde{W}_4^{12}$ to decide upon accepting or rejecting the offer. Calculations reveal that

$$\tilde{W}_4^{124} - \tilde{W}_4^{12} = \tilde{\Delta}_4 = 0.0623(a-c)^2 - 0.04428(a_4 - c)^2 \tag{31}$$

Once again, as shown in the Appendix 13, we can readily verify that the relationship between the choice of the ROW regarding accepting or rejecting the offer of accession and its market size relative to that of the member countries, as given by (31), is monotonically decreasing and the ROW will accept the offer of accession as long as its market size is less than the critical value $\overline{\overline{a'}} = (1.186a + 0.186c) > a$. Hence, for reasons absolutely similar to the case of Country-3 as discussed earlier, the ROW will also accept the offer of accession to form an enlarged FTA with countries 1 and 2 as long as its market size is less than a critical value $\overline{\overline{a'}} > a$.

Lemma 4 *Both the non-member countries will accept the offer of accession given their respective market sizes are less than a critical value* $\bar{\bar{a}}' = (1.186a + 0.186c)$.

5.2 *Simultaneous Consolidation and Expansion and the Non-member Country's Decision*

In this section, we consider the case of Country-3 being offered accession by the existing member countries to form an enlarged customs. Hence, the world system will look like [{1 + 2 + 3}, 4]. Under this scenario, the optimal tariff levels turn out to be

$$\hat{t}_{123} = \frac{7}{99}(2a + a_3 - 3c) \tag{32a}$$

$$\hat{t}_4^{123} = \frac{3}{11}(a_4 - c) = \tilde{t}_4^{12} \tag{32b}$$

Country-3 will decide upon accepting or rejecting the offer depending upon the sign of $\left(\hat{W}_3^{123} - \tilde{W}_3^{12}\right)$ and calculations[25] reveal that

$$\begin{aligned}(\hat{W}_3^{123} - \tilde{W}_3^{12}) &= \hat{\Delta}_3 \\ &= 0.058a^2 + 0.015aa_3 - 0.131ac - 0.044a_3^2 + 0.073a_3c + 0.029c^2\end{aligned} \tag{33}$$

[25]See Appendix 13.

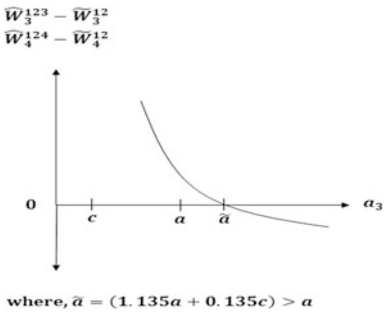


Fig. 5 Decision of the non-member countries

The sign of $\hat{\Delta}_3$ is ambiguous and hence, whether the national welfare of Country-3 is higher under CU_{123} regime or not depends, among others, on the magnitude of market size asymmetries of countries. The relationship as given by (33) is monotonically decreasing (shown in Appendix 14) and Country-3 will accept the offer of accession as long as its market size is less than the critical value $\tilde{a} = (1.135a + 0.135c) > a$. That is, Country-3 will accept the offer of accession to form an enlarged CU if the market size of Country-3 is less than the critical value $\tilde{a}$. This is illustrated in Fig. 5.

The economic intuition is exactly similar to that given in Sect. 4.1.1. Again, just like the calculations done in Sect. 4.1.1, here also we find that just like Country-3, the ROW will also accept the offer of accession to form an enlarged CU_{124} if $a_4 < \tilde{a}$.

Lemma 5 *Both the non-member countries will accept the offer of accession made by the initial member countries to form an enlarged customs union if their market sizes are less than a critical value* $\tilde{a} = (1.135a + 0.135c)$.

5.3 First-Stage Decisions

Given the optimal choices of the non-member countries when they are offered accession to FTA_{12}, as specified in Lemmas 4 and 5, we move on to analyze the optimal decisions of the existing member countries 1 and 2. We analyze the welfare levels of the initial member countries under all the possible alternative regimes. Once again, note that, the decisions of both the initial member countries will be exactly similar.[26]

To start with the welfare level of Country-1 under the benchmark FTA_{12} regime is

$$\tilde{W}_1^{12} = 0.4375(a-c)^2 + 0.008(a_3-c)^2 + 0.008(a_4-c)^2 \tag{34a}$$

[26] Hence, like Sect. 3, here also we will consider the decision of Country-1 with the understanding that decisions of both countries 1 and 2 will be the same.

Now, we proceed to analyze the welfare levels of Country-1 under the lateral expansion possibilities only (that is FTA$_{123}$ and FTA$_{124}$). Now, given FTA$_{123}$ regime, we have

$$\tilde{W}_1^{123} = 0.4116(a-c)^2 + 0.046(a_3-c)^2 + 0.008(a_4-c)^2 \tag{34b}$$

Subtracting (34a) from (34b) we have

$$\tilde{W}_1^{123} - \tilde{W}_1^{12} = 0.038(a_3-c)^2 - 0.026(a-c)^2 \tag{34c}$$

As shown in Appendix 15, it is readily verifiable that such relationship as given by (34c) is monotonically increasing and Country-1 will offer accession to Country-3 as long as the market size of the non-member Country-3 is greater than a critical level $a^* = (0.827a + 0.173c) < a$. Similarly if accession is offered to the ROW and it accepts, then under FTA$_{124}$ regime, using (63) and (34a) it can be easily verified that

$$\tilde{W}_1^{124} - \tilde{W}_1^{12} = 0.038(a_4-c)^2 - 0.026(a-c)^2 \tag{34d}$$

Hence, Country-1 will offer accession to the ROW as well if the market size of the ROW is larger than the same critical level a^*. This is because, the firm of Country-1 can use the non-member's market to sale the homogeneous good without any import tariff, and hence, its profit increases. Hence, larger the market size of the partner larger will be the profit from sales. Now, relevant calculations[27] reveal that

$$\tilde{W}_1^{123} - \tilde{W}_1^{124} = 0.038[(a_3-a_4)(a_3+a_4-2c)] > 0 \tag{34e}$$

That is, Country-1 will prefer the bigger non-member as a potential FTA partner.

Next, we analyze the welfare levels of Country-1 under the alternative simultaneous expansion possibilities (that is CU$_{123}$ and CU$_{124}$). Under CU$_{123}$ regime the welfare level of Country-1 turns out to be

$$\begin{aligned}\hat{W}_1^{123} =& 0.415a^2 + 0.012aa_3 - 0.842ac + 0.042a_3^2 - 0.097a_3c + 0.478c^2 \\ &+ 0.008a_4^2 - 0.016a_4c\end{aligned} \tag{35a}$$

Similarly, we can find out that

$$\begin{aligned}\hat{W}_1^{124} =& 0.415a^2 + 0.012aa_4 - 0.842ac + 0.042a_4^2 - 0.097a_4c + 0.478c^2 \\ &+ 0.008a_3^2 - 0.016a_3c\end{aligned} \tag{35b}$$

[27]See Appendix 16.

Comparing (35a) and (35b) we have

$$\hat{W}_1^{123} - \hat{W}_1^{124} = (a_3 - a_4)[0.012a + 0.034(a_3 + a_4) - 0.08c] > 0 \tag{35c}$$

We are still left with the possibility of the member countries opting for only consolidation without lateral expansion, that is, countries 1 and 2 decide only about consolidation of their economic cooperation by forming a customs union only between them (CU*12*). We obtain the welfare level of Country-1 under CU*12* as,

$$\hat{W}_1^{12} = 0.45(a-c)^2 + 0.008(a_3-c)^2 + 0.008(a_4-c)^2 \tag{36}$$

Now we start comparing the welfare levels of Country-1 for the alternative tariff regimes. Comparing (36) and (34a) we have

$$\hat{W}_1^{12} - \tilde{W}_1^{12} = 0.0125(a-c)^2 > 0 \tag{37}$$

That is, Country-1 will always prefer to form a CU with its existing partner Country-2 over an FTA with Country-2 irrespective of any parametric restriction on country sizes. Comparing (34b) and (36) we have

$$\hat{W}_1^{12} - \tilde{W}_1^{123} = 0.038(a-c)^2 - 0.038(a_3-c)^2 > 0 \quad \text{if } a > a_3 \tag{38}$$

Hence, if $a_3 > a$, then Country-1 will opt for an enlarged FTA rather than having a CU with Country-2 only. Similarly, we have

$$\hat{W}_1^{12} - \tilde{W}_1^{124} = 0.038(a-c)^2 - 0.038(a_4-c)^2 > 0 \quad \text{if } a > a_4 \tag{39}$$

However, from (34c) we already have $\tilde{W}_1^{123} > \tilde{W}_1^{124}$, that is, Country-1 will prefer the bigger non-member Country-3 as a partner to form the enlarged FTA.

Now, given the assumption that the initial member countries are symmetric with respect to both production technology and market size, Country-2 will face exactly the same conditions as that faced by Country-1 and hence, the decision of Country-1 is exactly the same that of Country-2.

Lemma 6 *Both the initial member countries will prefer to have an enlarged FTA and offer accession to the bigger non-member Country-3 as long as Country-3 has a bigger market size than the existing members.*

Comparing (35a) and (36) yields

$$\left(\hat{W}_1^{12} - \hat{W}_1^{123}\right) = 0.035a^2 - 0.012aa_3 - 0.057ac - 0.034a_3^2 + 0.08a_3c - 0.011c^2 \tag{40}$$

As shown in Appendix 17, the function $\left(\hat{W}_1^{12} - \hat{W}_1^{123}\right)$ is monotonically decreasing and we have $\hat{W}_1^{12} < \hat{W}_1^{123}$ for $a_3 > \bar{a} = (0.853a + 0.147c) < a$.

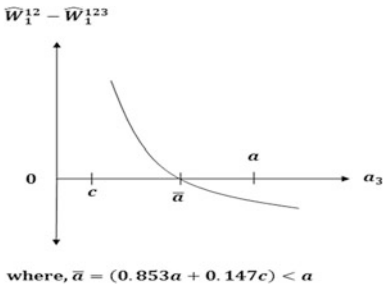


Fig. 6 Decision of the member countries

Similarly, for $a_3 < \bar{a}$ we have $\hat{W}_1^{12} > \hat{W}_1^{123}$. Note that the minimum value of a_3 for which positive quantities of the good will be produced is c. Now, as $a > c$, we have $\bar{a} < a$. Hence, Country-1 will offer find it profitable to offer accession to Country-3 in order to form an enlarged CU as long as $a_3 > \bar{a}$. However, from (35c) we already know that $\hat{W}_1^{123} > \hat{W}_1^{124}$. Hence, Country-1 will prefer Country-3 over the ROW to form an enlarged CU. This is illustrated in the following Fig. 6.

Once again, given exactly similar parameters, Country-2 will share the same view as Country-1 and hence, it will also offer accession to the bigger non-member Country-3 for the formation of an enlarged customs union.

Lemma 7 *Both the existing members will prefer to have an enlarged CU and offer accession to the bigger non-member Country-3 as long as Country-3 has a market size bigger than a critical limit in the sense that* $a_3 > \bar{a} = (0.853a + 0.147c)$.

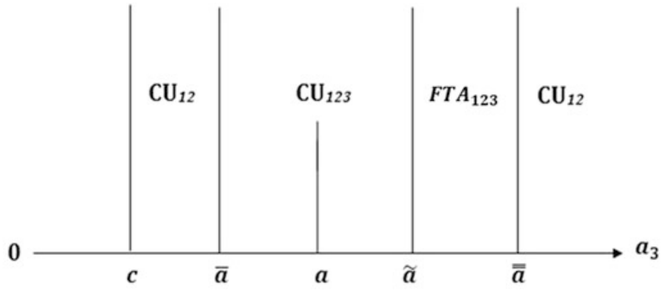
Now, we are in a position to define the SPNE of the game (illustrated in Fig. 1) under market size asymmetry.

5.4 *Subgame Perfect Expansion of the FTA*

Under the assumption of the countries being asymmetric only with respect to market size, we have two major cases to be considered—market size of Country-3 is smaller than that of the existing member countries $(a_3 < a)$ and market size of Country-3 is larger than that of the existing member countries $(a_3 > a)$.

However, given the additional assumption that Country-3 has a bigger market size than the ROW $(a_3 > a_4)$, we can unambiguously obtain $\tilde{W}_i^{123} > \tilde{W}_i^{124}$ and $\hat{W}_i^{123} > \hat{W}_i^{124}$ as given by (34c) and (35a) respectively. Again, irrespective of any assumptions regarding the ranking of a, a_3 and a_4, we have $\hat{W}_i^{12} > \tilde{W}_i^{12}\, \forall\, i = 1, 2$ from (37).

From the conditional welfare rankings obtained so far, we can reframe the two initial cases into a number of subcases depending on the relative market size of the bigger non-member country (Country-3). Hence, we need to establish the SPNE under each of the following subcases. From the detailed analyses (as shown in Appendix 18), we can establish the region-wise SPNE as follows (see Fig. 7 below).



$$\bar{a} = (0.853a + 0.147c), \tilde{a} = (1.135a + 0.135c), \bar{\bar{a}} = (1.187a + 0.187c)$$

Fig. 7 Different market sizes of country-3 and SPNE

Region I $[c < a_3 < \bar{a}]$: We have formation of CU_{12} as the unique SPNE in this region.

Region II $[\bar{a} < a_3 < a]$: In this region have formation of CU_{123} as the unique SPNE given $a_3 > a_4$. If $a_3 < a_4$, then formation of CU_{124} would have been the unique SPNE.

Region III $[a < a_3 < \tilde{a}]$: Same as Region II.

Region IV $[\tilde{a} < a_3 < \bar{\bar{a}}]$: In this region FTA_{123} turns out to be the unique SPNE.

Region V $[a_3 > \bar{\bar{a}}]$: Same as Region I.

Proposition 2 *There exist multiple SPNE under market size asymmetry among the non-member countries. The outcome depends on the market size of the bigger non-member country.*

Proof Follows from Appendix 18.

Thus there exists no unique SPNE to the game (illustrated in Fig. 1) given market size asymmetry among the non-member countries. The outcomes are dependent on the relative size of the bigger non-member country. However, it is more likely that there will be deeper cooperation between initial member countries with or without simultaneous inclusion of a new member.

6 Conclusion

This paper analyzes the conditions under which an existing RTB prefers an expansion and when it prefers consolidation through higher order trade blocs or both simultaneously. In doing so it tries to link two related streams of literature on regional economic integration—the set that analyzes equilibrium structure of RTBs and that analyzing the effects of policy centralization on the relation between widening and deepening choices faced by existing unions. The analysis has been done assuming an asymmetric world, both with respect to production technology and domestic market sizes. The preference of the existing members for a non-member in case the existing RTB opts for expansion has also been addressed.

The benchmark analysis suggests that a customs union between two similar countries will be sustained as the SPNE under unanimous regionalism under the assumption that the initial non-member countries differ only in respect of production technology. However, given market size asymmetry, the SPNE outcome depends completely on the market size of the larger non-member country. Customs union among the initial member countries can be sustained as the SPNE for smallest and largest critical levels of the market size of the larger non-member country. For the intermediate values both FTA and CU with the larger non-member country can be sustained as the SPNE outcome.

These results imply that under technology and market size asymmetries, existing blocs tend to expand in different patterns depending on the nature and extent of asymmetries among the potential union members. This observation is true, if we have a closer look at the present scenario of the world trading system—not only new blocs have came up but old blocs have expanded with different degrees of economic cooperation among the member countries.

However, this model can be extended further to incorporate the possibility of simultaneous formation of multiple trading blocs.

Acknowledgments The author deeply acknowledges the theoretical inputs from Rajat Acharyya. Suggestions from Brati Sankar Chakraborty, Kalyan Sanyal, Saibal Kar, Sarbajit Sengupta, Soumyendranath Sikdar and Sugata Marjit on an earlier draft of the paper are also acknowledged. The author also thanks the referee for helpful comments. The usual disclaimer is, however, applied.

Appendix

1. Optimal Tariffs under Benchmark Case

From (4) we have the welfare function for Country-1 to be

$$W_1 = \pi_1 + \mathrm{CS}_1 + \mathrm{TR}_1$$

Under FTA$_{12}$ regime using (3) we have

$$\pi_1 = P_1X_{11} + P_2X_{12} + P_3X_{13} + P_4X_{14} - c_1(X_{11} + X_{12} + X_{13} + X_{14}) - t_3X_{13} - t_4X_{14}$$

Now, $\mathrm{CS}_1 = \frac{1}{2}(a - P_1)^2$ and $\mathrm{TR}_1 = t_1(X_{31} + X_{41})$

Table 1 Trading arrangements of the emerging economies

Countries	CU, CU & EIA	FTA, FTA & EIA	PSA
Argentina	MERCOSUR		GSTP, LAIA, MERCOSUR-India
Brazil	MERCOSUR		GSTP, LAIA, MERCOSUR-India
Bulgaria	EU		
Chile			GSTP, LAIA, India
China		ASEAN, Chile, Costa Rica, Hong Kong, Macao, NZ, Singapore, Peru, Pakistan, Switzerland, Iceland	APTA
Colombia	CAN	Canada, Chile, Mexico, Northern Triangle (El Salvador, Guatemala, Honduras), EFTA (Iceland, Liechtenstein, Norway, Switzerland), EU, Peru, USA	GSTP, LAIA
Estonia	EU		
Hungary	EU		
India		ASEAN, Bhutan, Sri Lanka, SAFTA, South Korea, Singapore, Malaysia, Japan	Afghanistan, Chile, Nepal, MERCOSUR, SAPTA, APTA, GSTP
Indonesia		ASEAN, Japan	GSTP
Latvia	EU		
Lithuania	EU		
Malaysia		ASEAN, AFTA, Chile, Pakistan, NZ, Australia, Japan, India	GSTP
Mexico		NAFTA, Chile, Colombia, EFTA (Iceland, Liechtenstein, Norway, Switzerland), EU, Japan, Central America (Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua), Uruguay, Peru, Israel	GSTP, LAIA, PTN
Pakistan		China, Malaysia, SAFTA, Sri Lanka	SAPTA, ECO (Iran, Turkey), GSTP, PTN
Peru	CAN	Canada, Costa Rica, EU, Colombia, Japan, Panama, Chile, China, South Korea, Mexico, Singapore, USA, EFTA	GSTP, PTN, LAIA
Philippines		ASEAN, Japan	GSTP, PTN
Poland	EU		
Romania	EU		

(continued)

Table 1 (continued)

Countries	CU, CU & EIA	FTA, FTA & EIA	PSA
Russia	EAEC, Belarus–Kazakhstan	Armenia, CEZ, Georgia, Kyrgyz, Azerbaijan, Moldova, Serbia, Tajikistan, Turkmenistan, Uzbekistan, Ukraine, CIS, EAEU	
South Africa	SACU	SADC, EU, EFTA-SACU	
Thailand		ASEAN, Japan, Australia, NZ	Laos, GSTP
Turkey	EU	South Korea, Albania, B&H, EFTA, Egypt, Chile, Macedonia, Israel, Georgia, Jordan, Mauritius, Morocco, Montenegro, Palestine, Serbia, Tunisia, Syria	ECO, PTN
Ukraine		CEZ, Armenia, Georgia, Kyrgyz, CIS, Belarus, Azerbaijan, Kazakhstan, Moldova, Russia, Uzbekistan, Tajikistan, EFTA, Montenegro	
Venezuela	CAN		GSTP, LAIA

Source Author's compilation from WTO RTA Database

Therefore we have

$$W_1 = P_1X_{11} + P_2X_{12} + P_3X_{13} + P_4X_{14} - c_1(X_{11} + X_{12} + X_{13} + X_{14}) - t_3X_{13} - t_4X_{14} + \frac{1}{2}(a - P_1)^2 + t_1(X_{31} + X_{41}) \tag{41}$$

Putting the values of P_i and X_{ij} from (5a–5g) and (6a–6b) in (41) and with respect to t_1 yields optimal $\tilde{t}_1^{12}$. Proceeding similar for countries 2, 3, and ROW we have the optimal tariffs as given in (7a–7b).

2. Welfare Comparison for Country-3

From (4) we have the welfare function for Country-3 to be

$$W_3 = \pi_3 + \mathrm{CS}_3 + \mathrm{TR}_3$$

Under FTA$_{12}$ regime using (3) we have

$$\pi_3 = P_1X_{31} + P_2X_{32} + P_3X_{33} + P_4X_{34} - c_3(X_{31} + X_{32} + X_{33} + X_{34}) - t_1X_{31} - t_2X_{32} - t_4X_{34}$$

Now, $CS_3 = \frac{1}{2}(a - P_3)^2$ and $TR_3 = t_3(X_{13} + X_{23} + X_{43})$

$$\begin{aligned}\therefore W_3 = {} & P_1X_{31} + P_2X_{32} + P_3X_{33} + P_4X_{34} - c_3(X_{31} + X_{32} + X_{33} + X_{34}) - t_1X_{31} - t_2X_{32} \\ & - t_4X_{34} + \frac{1}{2}(a - P_3)^2 + t_3(X_{13} + X_{23} + X_{43})\end{aligned} \tag{42}$$

Putting equilibrium tariffs $\tilde{t}_i^{12}$ in (5a–5g) and (6a–6b) we get the equilibrium output values ($\tilde{X}_{ij}^{12}$) and equilibrium prices ($\tilde{P}_i^{12}$). Putting those values in (42) we get

$$\begin{aligned}\tilde{W}_3^{12} = {} & 0.448a^2 + 0.02ac - 1.033ac_3 + 0.115ac_4 + 0.547c^2 - 1.902cc_3 \\ & + 0.786cc_4 + 2.248c_3^2 - 1.562c_3c_4 + 0.33c_4^2\end{aligned} \tag{43}$$

For FTA$_{123}$ regime, the welfare function for Country-3 is given by

$$\begin{aligned}W_3 = {} & P_1X_{31} + P_2X_{32} + P_3X_{33} + P_4X_{34} - c_3(X_{31} + X_{32} + X_{33} + X_{34}) - t_4X_{34} \\ & + \frac{1}{2}(a - P_3)^2 + t_3X_{43}\end{aligned} \tag{44}$$

Putting equilibrium tariffs $\tilde{t}_i^{123}$ in output and price functions we get the equilibrium output values ($\tilde{X}_{ij}^{123}$) and equilibrium prices ($\tilde{P}_i^{123}$). Putting those values in (44) we get

$$\begin{aligned}\tilde{W}_3^{123} = {} & 0.467a^2 + 0.343ac - 1.284ac_3 + 0.008ac_4 + 0.932c^2 - 2.572cc_3 \\ & + 0.364cc_4 + 2.428c_3^2 - 1.000c_3c_4 + 0.314c_4^2\end{aligned} \tag{45}$$

Hence, subtracting (43) from (45) we have

$$\begin{aligned}\tilde{W}_3^{123} - \tilde{W}_3^{12} = {} & 0.018a^2 + 0.323ac - 0.252ac_3 - 0.107ac_4 + 0.385c^2 - 0.671cc_3 \\ & - 0.422cc_4 + 0.18c_3^2 + 0.562c_3c_4 - 0.016c_4^2\end{aligned}$$

3. Decision of Country-3 regarding Joining FTA$_{123}$

From (10) we have: $\tilde{\Delta}_3 = 0.018a^2 + 0.323ac - 0.359ac_3 + 0.385c^2 - 1.093cc_3 + 0.726c_3^2$

$$\therefore \frac{\delta}{\delta c_3}(\tilde{\Delta}_3) = -0.359a - 1.093c + 1.452c_3 = -0.359(a - c_3) + 1.093(c_3 - c) \tag{46a}$$

and

$$\frac{\delta^2}{\delta c_3^2}(\tilde{\Delta}_3) = 1.452 > 0 \tag{46b}$$

Moreover, it can be easily calculated that $\tilde{\Delta}_3 = 0$ for $c_3 = \tilde{c}, \tilde{\tilde{c}}$ where

$$c < \tilde{\tilde{c}} = (0.056a + 0.944c) < \tilde{c} = (0.438a + 0.562c) < a \tag{47a}$$

Now, if we assume that $c_3 = c_4 = c$, then we have

$$\tilde{\Delta}_3 = 0.018(a - c)^2 > 0 \tag{47b}$$

Again, if $c_3 = c_4 = a$, then we have

$$\tilde{\Delta}_3 = 0.385(a - c)^2 > 0 \tag{47c}$$

Hence, using (10) and (46a)–(47c) we can easily claim that Country-3 will accept the offer of accession to the existing FTA$_{12}$ in the range $(c, \tilde{\tilde{c}})$ and $(\tilde{c}, a)$.

4. Joint Welfare Maximization and Common External Tariff

Let the joint welfare of all the member countries under CU$_{123}$ be

$$\begin{aligned}\hat{W}_{123}^{123} &= \hat{W}_1^{123} + \hat{W}_2^{123} + \hat{W}_3^{123} \\ &= P_1X_{11} + P_2X_{12} + P_3X_{13} + P_4X_{14} - c_1(X_{11} + X_{12} + X_{13} + X_{14}) - t_4X_{14} \\ &\quad + P_1X_{21} + P_2X_{22} + P_3X_{23} + P_4X_{24} - c_2(X_{21} + X_{22} + X_{23} + X_{24}) - t_4X_{24} \\ &\quad + P_1X_{31} + P_2X_{32} + P_3X_{33} + P_4X_{34} - c_3(X_{31} + X_{32} + X_{33} + X_{34}) - t_4X_{34} \\ &\quad + \frac{1}{2}(a - P_1)^2 + \frac{1}{2}(a - P_2)^2 + \frac{1}{2}(a - P_3)^2 + \hat{t}(X_{41} + X_{42} + X_{43})\end{aligned} \tag{48}$$

Now, putting the optimal values of $\hat{X}_{ij}$ and $\hat{P}_i$ in (48) and maximizing with respect to $\hat{t}$ yields the optimal CET $\hat{t}_{123}$.

5. Welfare Comparisons for Country-3 under Alternative Regimes

The welfare of Country-3 under CU$_{123}$ regime is given to be

$$\hat{W}_3^{123} = P_1X_{31} + P_2X_{32} + P_3X_{33} + P_4X_{34} - c_3(X_{31} + X_{32} + X_{33} + X_{34}) - t_4X_{34} + \frac{1}{2}(a - P_3)^2 + \hat{t}X_{43} \tag{49}$$

Now, putting the optimal tariffs from (15a)–(15b) in (49) we have

$$\hat{W}_3^{123} = 0.478a^2 + 0.441ac - 1.416ac_3 + 0.0196ac_4 + 0.847c^2 - 2.547cc_3 + 0.413cc_4 + 2.515c_3^2 - 1.066c_3c_4 + 0.317c_4^2 \tag{50}$$

Now, subtracting (43) from (50) we get

$$\hat{W}_3^{123} - \tilde{W}_3^{12} = 0.03a^2 + 0.421ac - 0.383ac_3 - 0.096ac_4 + 0.3c^2 - 0.645cc_3 - 0.373cc_4 + 0.267c_3^2 + 0.496c_3c_4 - 0.013c_4^2$$

6. Decision of Country-3 regarding Joining CU$_{123}$

From (17) we have

$$\hat{\Delta}_3 = 0.03a^2 + 0.421ac - 0.479ac_3 + 0.3c^2 - 1.018cc_3 + 0.75c_3^2$$

$$\begin{aligned}\therefore \frac{\delta}{\delta c_3}(\hat{\Delta}_3) &= -0.479a - 1.018c + 1.5c_3\\ &= -0.479(a - c_3) + 1.018(c_3 - c) + 0.003c_3\end{aligned}$$

and

$$\frac{\delta^2}{\delta c_3^2}(\hat{\Delta}_3) = 1.5 > 0 \tag{51}$$

Moreover, it can be easily calculated that $\hat{\Delta}_3 = 0$ for $c_3 = \hat{c}, \hat{\hat{c}}$ where

$$c < \hat{\hat{c}} = (0.07a + 0.93c) < \hat{c} = (0.57a + 0.43c) < a \tag{52a}$$

Now, if we assume that $c_3 = c_4 = c$, then we have

$$\hat{\Delta}_3 = 0.03(a - c)^2 > 0 \tag{52b}$$

Again, if $c_3 = c_4 = a$, then we have

$$\hat{\Delta}_3 = 0.3(a - c)^2 > 0 \tag{52c}$$

Hence, using (17) and (51)–(52c) we can easily claim that Country-3 will accept the offer of accession to the new CU$_{123}$ in the range $(c, \hat{\hat{c}})$ and $(\hat{c}, a)$.

7. Welfare Levels of Country-1 under Alternative Regimes

Using the optimal tariff values given by (8a)–(8c) in the welfare function of Country-1 under the FTA$_{123}$ regime we have:

$$\tilde{W}_1^{123} = 0.467a^2 - 1.113ac + 0.171ac_3 + 0.008ac_4 + 1.367c^2 - 0.823cc_3 - 0.818cc_4 + 0.234c_3^2 + 0.182c_3c_4 + 0.314c_4^2$$

Similarly using the optimal tariff values given by (11a)–(11c) we have

$$\tilde{W}_1^{124} = 0.467a^2 - 1.113ac + 0.008ac_3 + 0.171ac_4 + 1.367c^2 - 0.818cc_3 - 0.823cc_4 + 0.314c_3^2 + 0.182c_3c_4 + 0.234c_4^2$$

8. Welfare of Country-1 under CU$_{123}$ and CU$_{124}$

The welfare of Country-1 under CU$_{123}$ regime is given to be

$$\hat{W}_1^{123} = P_1X_{11} + P_2X_{12} + P_3X_{13} + P_4X_{14} - c(X_{11} + X_{12} + X_{13} + X_{14}) - t_4X_{14} + \frac{1}{2}(a - P_1)^2 + \hat{t}X_{41} \tag{53}$$

Now, putting the optimal tariffs from (15a)–(15b) in (53) we have

$$\hat{W}_1^{123} = 0.478a^2 - 1.196ac + 0.22ac_3 + 0.019ac_4 + 1.453c^2 - 0.85cc_3 - 0.859cc_4 + 0.217c_3^2 + 0.207c_3c_4 + 0.317c_4^2$$

The welfare of Country-1 under CU$_{124}$ regime is given to be

$$\hat{W}_1^{124} = P_1X_{11} + P_2X_{12} + P_3X_{13} + P_4X_{14} - c(X_{11} + X_{12} + X_{13} + X_{14}) - t_3X_{13} + \frac{1}{2}(a - P_1)^2 + \hat{t}X_{31} \tag{54}$$

Now, putting the optimal tariffs from (18a)–(18b) in (54) we have

$$\hat{W}_1^{124} = 0.478a^2 - 1.196ac + 0.019ac_3 + 0.22ac_4 + 1.453c^2 - 0.859cc_3 - 0.85cc_4 + 0.317c_3^2 + 0.207c_3c_4 + 0.217c_4^2$$

9. Welfare Rankings for Country-1

Comparing (20) and (27) we get

$$\begin{aligned}\hat{W}_1^{12} - \tilde{W}_1^{12} &= 0.144a^2 + 0.093ac + 0.189ac_3 + 0.189ac_4 - 0.606c^2 + 0.474cc_3 \\ &\quad + 0.474cc_4 + 0.004c_3^2 + 0.006c_3c_4 + 0.004c_4^2 \\ &= 0.144a^2 + 0.093ac + 0.189ac_3 + 0.189ac_4 + 0.474c(c_3 - c) \\ &\quad + 0.132c(c_4 - c) + 0.342cc_4 + 0.004c_3^2 + 0.006c_3c_4 + 0.004c_4^2 > 0\end{aligned}$$

[$\because$ by assumption $c_3 > c_4 > c$]
Comparing (21) and (27) we have

$$\begin{aligned}\hat{W}_1^{12} - \tilde{W}_1^{123} &= 0.131a^2 + 0.247ac + 0.043ac_3 + 0.206ac_4 - 0.56c^2 + 0.364cc_3 \\ &\quad + 0.359cc_4 - 0.002c_3^2 + 0.274c_3c_4 - 0.082c_4^2 \\ &= 0.129a^2 + 0.002(a^2 - c_3^2) + 0.247ac + 0.043ac_3 + 0.206ac_4 \\ &\quad + 0.364c(c_3 - c) + 0.196c(c_4 - c) + 0.163cc_4 + 0.192c_3c_4 \\ &\quad + 0.082c_4(c_3 - c_4) > 0\end{aligned}$$

[$\because$ by assumption $a > c_3 > c_4 > c$]
From (22) and (27) we have

$$\begin{aligned}\hat{W}_1^{12} - \tilde{W}_1^{124} &= 0.131a^2 + 0.247ac + 0.206ac_3 + 0.043ac_4 - 0.56c^2 + 0.359cc_3 \\ &\quad + 0.364cc_4 - 0.082c_3^2 + 0.274c_3c_4 - 0.002c_4^2 \\ &= 0.049a^2 + 0.082(a^2 - c_3^2) + 0.247ac + 0.206ac_3 + 0.043ac_4 \\ &\quad + 0.359c(c_3 - c) + 0.201c(c_4 - c) + 0.163cc_4 + 0.272c_3c_4 \\ &\quad + 0.002c_4(c_3 - c_4) > 0\end{aligned}$$

[$\because$ by assumption $a > c_3 > c_4 > c$]
Subtracting (23) from (27) we have

$$\begin{aligned}\hat{W}_1^{12} - \hat{W}_1^{123} &= 0.12a^2 + 0.33ac - 0.006ac_3 + 0.195ac_4 - 0.646c^2 + 0.391cc_3 \\ &\quad + 0.4cc_4 + 0.015c_3^2 + 0.249c_3c_4 - 0.085c_4^2 \\ &= 0.114a^2 + 0.006a(a - c_3) + 0.33ac + 0.195ac_4 + 0.391c(c_3 - c) \\ &\quad + 0.255c(c_4 - c) + 0.145cc_4 + 0.015c_3^2 + 0.164c_3c_4 \\ &\quad + 0.085(c_3 - c_4) > 0\end{aligned}$$

[$\because$ by assumption $a > c_3 > c_4 > c$]

Similarly from (24) and (27) we get

$$\begin{aligned}\hat{W}_1^{12} - \tilde{W}_1^{124} &= 0.12a^2 + 0.33ac + 0.195ac_3 - 0.006ac_4 - 0.646c^2 + 0.4cc_3 \\ &\quad + 0.391cc_4 - 0.085c_3^2 + 0.249c_3c_4 + 0.015c_4^2 \\ &= 0.029a^2 + 0.085\left(a^2 - c_3^2\right) + 0.33ac + 0.195ac_3 + 0.006a(a - c_4) \\ &\quad + 0.4c(c_3 - c) + 0.246c(c_4 - c) + 0.145cc_4 + 0.249c_3c_4 + 0.015c_4^2 > 0\end{aligned}$$

[$\because$ by assumption $a > c_3 > c_4 > c$]

10. Optimal Tariffs under Market Size Asymmetry

Welfare function for Country-1 is given by

$$\begin{aligned}\tilde{W}_1^{12} &= P_1X_{11} + P_2X_{12} + P_3X_{13} + P_4X_{14} - c_1(X_{11} + X_{12} + X_{13} + X_{14}) - t_3X_{13} - t_4X_{14} \\ &\quad + \frac{1}{2}(a - P_1)^2 + t_1(X_{31} + X_{41}) = \frac{1}{25}(a + 4c + 2t_1)(a - c + 2t_1) + \frac{1}{25}(a + 4c \\ &\quad + 2t_2)(a - c + 2t_2) + \frac{1}{25}(a_3 + 4c + 3t_3)(a_3 - c - 2t_3) + \frac{1}{25}(a_4 + 4c \\ &\quad + 3t_4)(a_4 - c - 2t_4) - \frac{1}{5}c(2a + a_3 + a_4 - 4c + 2t_1 + 2t_2 - 2t_3 - 2t_4) - t_3X_{13} \\ &\quad - t_4X_{14} + \frac{1}{50}(4a - 4c - 2t_1)^2 + \frac{2}{5}(a - c - 3t_1)t_1\end{aligned} \tag{55}$$

Maximizing (55) with respect to t_1 yields optimal tariff $\tilde{t}_1^{12} = \frac{1}{8}(a - c)$. Proceeding in the same fashion yields the optimal tariffs as given in (28a) and (28b).

11. Welfare Comparisons for Country-3

The welfare function of Country-3 under FTA$_{12}$ regime can be written as

$$\begin{aligned}W_3 &= P_1X_{31} + P_2X_{32} + P_3X_{33} + P_4X_{34} - c_3(X_{31} + X_{32} + X_{33} + X_{34}) - t_1X_{31} - t_2X_{32} \\ &\quad - t_4X_{34} + \frac{1}{2}(a - P_3)^2 + t_3(X_{13} + X_{23} + X_{43})\end{aligned}$$

Putting equilibrium tariffs levels $\tilde{t}_1^{12}$ from (28a and 28b), the equilibrium output values $(\tilde{X}_{ij}^{12})$ and prices $\tilde{P}_i^{12}$ in the above equation we have

$$\tilde{W}_3^{12} = 0.03125(a - c)^2 + 0.4091(a_3 - c)^2 + 0.008(a_4 - c)^2 \tag{56}$$

The welfare function of Country-3 under FTA$_{123}$ regime can be written as

$$\begin{aligned}W_3 &= P_1X_{31} + P_2X_{32} + P_3X_{33} + P_4X_{34} - c_3(X_{31} + X_{32} + X_{33} + X_{34}) - t_4X_{34} \\ &\quad + \frac{1}{2}(a - P_3)^2 + t_3X_{43}\end{aligned}$$

Putting equilibrium tariffs levels $\tilde{t}_i^{123}$ from (29a–29c), the equilibrium output values $\left(\tilde{X}_{ij}^{123}\right)$ and prices $\tilde{P}_i^{123}$ in the above equation we have

$$\tilde{W}_3^{123} = 0.0935(a-c)^2 + 0.3648(a_3-c)^2 + 0.008(a_4-c)^2 \tag{57}$$

Hence, subtracting (56) from (57) we have

$$\tilde{W}_3^{123} - \tilde{W}_3^{12} = 0.062a^2 - 0.044a_3^2 + 0.018c^2 + 0.088a_3c - 0.124ac$$

12. Decision of Country-3 in Joining FTA$_{123}$ under Market Size Asymmetry

From (30) we have:

$$\tilde{W}_3^{123} - \tilde{W}_3^{12} = 0.062a^2 - 0.044a_3^2 + 0.018c^2 + 0.088a_3c - 0.124ac$$

$$\therefore \quad \frac{\partial}{\partial a_3}\left(\tilde{W}_3^{123} - \tilde{W}_3^{12}\right) = -0.088a_3 + 0.088c < 0 \tag{58a}$$

$$\frac{\partial^2}{\partial a_3^2}\left(\tilde{W}_3^{123} - \tilde{W}_3^{12}\right) = -0.088 < 0 \tag{58b}$$

$$\text{Further we have } \left(\tilde{W}_3^{123} - \tilde{W}_3^{12}\right) = 0 \text{ if } a_3 = (1.187a + 0.187c) = \bar{\bar{a}} > a \tag{58c}$$

$$\text{Again if } a_3 = a, \text{ then } \left(\tilde{W}_3^{123} - \tilde{W}_3^{12}\right) = 0.018(a-c)^2 > 0 \tag{58d}$$

Hence, Country-3 will accept the offer of accession to form an enlarged FTA only if $a_3 < \bar{\bar{a}}$.

13. Decision of the ROW in Joining FTA$_{124}$ under Market Size Asymmetry

From (31) it is easily evident that $\left(\tilde{W}_4^{124} - \tilde{W}_4^{12}\right)$ is a monotonically decreasing function.

$$\frac{\partial}{\partial a_3}\left(\tilde{W}_4^{124} - \tilde{W}_4^{12}\right) = -0.08856(a_4 - c) < 0 \tag{59a}$$

$$\frac{\partial^2}{\partial a_3^2}\left(\tilde{W}_4^{124} - \tilde{W}_4^{12}\right) = -0.08856 < 0 \tag{59b}$$

$$\text{Further we have } \left(\tilde{W}_4^{124} - \tilde{W}_4^{12}\right) = 0 \text{ if } a_4 = (1.186a + 0.186c) = \overline{\overline{a'}} > a \tag{59c}$$

$$\text{Again if } a_4 = a, \text{then } \left(\tilde{W}_4^{124} - \tilde{W}_4^{12}\right) = 0.01802(a-c)^2 > 0 \tag{59d}$$

Hence, the ROW will accept the offer of accession as long as $a_4 < \overline{\overline{a'}}$.

14. Welfare Comparison and Decision of Country-3 in Joining CU_{123}

Using the equilibrium output $\hat{X}_{ij}^{123}$ and prices $\hat{P}_i^{123}$ in (49) we have

$$\begin{aligned}\hat{W}_3^{123} &= 0.089a^2 + 0.015aa_3 - 0.194ac + 0.0365a_3^2 - 0.745a_3c \\ &\quad + 0.478c^2 + 0.008a_4^2 - 0.016a_4c\end{aligned} \tag{60}$$

Subtracting (56) from (60) we have

$$\left(\hat{W}_3^{123} - \tilde{W}_3^{12}\right) = 0.058a^2 + 0.015aa_3 - 0.131ac - 0.044a_3^2 + 0.073a_3c + 0.029c^2$$

Now we perform the following exercises.

$$\frac{\partial}{\partial a_3}\left(\hat{W}_3^{123} - \tilde{W}_3^{12}\right) = 0.015a - 0.088a_3 + 0.073c \tag{61a}$$

$$\frac{\partial^2}{\partial a_3^2}\left(\hat{W}_3^{123} - \tilde{W}_3^{12}\right) = -0.088 < 0 \tag{61b}$$

$$\text{Now, } \left(\hat{W}_3^{123} - \tilde{W}_3^{12}\right) = 0 \text{ if } a_3 = (1.135a + 0.135c) = \tilde{a} > a \tag{61c}$$

$$\text{Again if } a_3 = a, \text{then } \left(\hat{W}_3^{123} - \tilde{W}_3^{12}\right) = 0.029(a-c)^2 > 0 \tag{61d}$$

Combining the results from (61a)–(61d) we have

$$(\hat{W}_3^{123} - \tilde{W}_3^{12}) \text{ if } a_3 < \tilde{a}$$

That is, Country-3 will accept the offer of accession to form an enlarged CU if the market size of Country-3 is less than a critical value $\tilde{a} = (1.135a + 0.135c)$.

15. Lateral Expansion and Decision of Country-1

From (34c) it is easily evident that $\left(\tilde{W}_1^{123} - \tilde{W}_1^{12}\right)$ is a monotonically increasing function.

$$\frac{\partial}{\partial a_3}\left(\tilde{W}_1^{123} - \tilde{W}_1^{12}\right) = 0.076(a_3 - c) > 0 \tag{62a}$$

$$\frac{\partial^2}{\partial a_3^2}\left(\tilde{W}_1^{123} - \tilde{W}_1^{12}\right) = 0.076 > 0 \tag{62b}$$

$$\text{Further we have } \left(\tilde{W}_1^{123} - \tilde{W}_1^{12}\right) = 0 \text{ if } a_3 = (0.827a + 0.173c) = a^* < a \quad (62c)$$

$$\text{Again if } a_3 = a, \text{then } \left(\tilde{W}_1^{123} - \tilde{W}_1^{12}\right) = 0.012(a-c)^2 > 0 \quad (62d)$$

Hence, Country-1 will offer accession to Country-3 if $a_3 > a^*$.

16. Preference of Non-member by Country-1 for Lateral Expansion

From (34b) we have:

$$\tilde{W}_1^{123} = 0.4116(a-c)^2 + 0.046(a_3-c)^2 + 0.008(a_4-c)^2$$

Similarly, for Country-1 under FTA$_{124}$ regime we have

$$\tilde{W}_1^{124} = 0.4116(a-c)^2 + 0.008(a_3-c)^2 + 0.046(a_4-c)^2 \quad (63)$$

Hence, subtracting (63) from (34b) we have

$$\tilde{W}_1^{123} - \tilde{W}_1^{124} = 0.038[(a_3-a_4)(a_3+a_4-2c)] > 0$$

17. Simultaneous Expansion and Choice of Country-1

The function $(\hat{W}_1^{12} - \hat{W}_1^{123})$ is monotonically decreasing.

$$\frac{\partial}{\partial a_3}(\varDelta\hat{W}_1) = -[0.012(a-c) + 0.068(a_3-c)] < 0 \quad (64a)$$

$$\frac{\partial^2}{\partial a_3^2}(\varDelta\hat{W}_1) = -0.068 < 0 \quad (64b)$$

Further we have, if $a = a_3$ then we have

$$\varDelta\hat{W}_1 = -0.014(a-c)^2 < 0 \quad (64c)$$

Additional calculations reveal that

$$\varDelta\hat{W}_1 = 0 \text{ if } a_3 = \bar{a} \text{ where } \bar{a} = (0.853a + 0.147c) < a \quad (64d)$$

Therefore, from (64a)–(64d) we have

$$\hat{W}_1^{12} < \hat{W}_1^{123} \text{ for } a_3 > \bar{a} \text{ and} \quad (65a)$$

$$\hat{W}_1^{12} > \hat{W}_1^{123} \text{ for } a_3 < \bar{a} \quad (65b)$$

18. Proof of Proposition 2

Region I: $c < a_3 < \bar{a}$

In Region I, the bigger non-member country (Country-3) will accept the offer of accession by the initial FTA$_{12}$ members for both FTA and CU regimes as in this region we have

$$\hat{W}_3^{123} > \tilde{W}_3^{12} \quad \text{[from Appendix 14]}$$

$$\tilde{W}_3^{123} > \tilde{W}_3^{12} \quad \text{[from Appendix 12]}$$

However, for both the initial members (Country-1 and 2), it is better to opt for deeper integration among themselves rather than expanding membership of the existing trade bloc. This is because, in this region, the following welfare rankings hold.

$$\hat{W}_i^{12} > \tilde{W}_i^{12} \quad \forall \quad i = 1, 2. \quad \text{[from (37)]}$$

$$\hat{W}_i^{12} > \hat{W}_i^{123} \quad \forall \quad i = 1, 2. \quad \text{[from Appendix 14]}$$

$$\hat{W}_i^{12} > \tilde{W}_i^{123} \quad \forall \quad i = 1, 2. \quad \text{[from (38)]}$$

Again, given $a_3 > a_4$ we already have the following rankings

$$\hat{W}_i^{123} > \hat{W}_i^{124} \quad \forall \quad i = 1, 2.$$

$$\tilde{W}_i^{123} > \tilde{W}_i^{124} \quad \forall \quad i = 1, 2.$$

Hence, using the method of backward induction, we have **CU$_{12}$** as the unique SPNE in this region.

Region II: $\bar{a} < a_3 < a$

Already, we know that irrespective of the size of the bigger non-member Country-3

$$\hat{W}_i^{12} > \tilde{W}_i^{12} \quad \forall \quad i = 1, 2.$$

In this region as $a_3 > \bar{a}$, we have

$$\hat{W}_i^{123} > \hat{W}_i^{12} \quad \forall \quad i = 1, 2.$$

Further, we have $a_3 < a$

$$\hat{W}_i^{12} > \tilde{W}_i^{123} \quad \forall \quad i = 1, 2.$$

Combining all these we have

$$\hat{W}_i^{123} > \hat{W}_i^{12} > \tilde{W}_i^{123} > \tilde{W}_i^{12} \quad \forall \quad i = 1, 2.$$

Now, as long as $a_3 < \tilde{a}$ we have $\hat{W}_3^{123} > \tilde{W}_3^{12}$ and given $\tilde{a} > a$, we definitely know that in this region the non-member Country-3 will accept the offer of accession to form a CU with the initial member countries.

Hence in this region, we have formation of $\mathbf{CU}_{\mathbf{123}}$ as the unique SPNE. However, if $a_3 < a_4$, that is if Country-3 had a smaller market size than Country-4, then $\mathbf{CU}_{\mathbf{124}}$ would have been the unique SPNE instead of $\mathbf{CU}_{\mathbf{123}}$.

Region III: $\boldsymbol{a < a_3 < \tilde{a}}$

In this region as $a_3 > a$, we have

$$\tilde{W}_i^{123} > \hat{W}_i^{12} \quad \forall \quad i = 1, 2.$$

Given $a_3 > \bar{a}$, we also have

$$\hat{W}_i^{123} > \hat{W}_i^{12} \quad \forall \quad i = 1, 2.$$

Now, the whether the existing FTA$_{12}$ members will choose to expand (to FTA$_{123}$) or expand and integrate at the same time (CU$_{123}$) depends on the relative strength of the welfare levels under the concerned alternative regimes.

$$\hat{W}_i^{123} - \tilde{W}_i^{123} = 0.003a^2 + 0.012aa_3 - 0.019ac - 0.005a_3^2 - 0.004a_3c + 0.011c^2$$

Now from the above equation it is not possible to unambiguously determine the exact ranking of $\hat{W}_i^{123}$ and $\tilde{W}_i^{123}$. However, $\hat{W}_i^{123} - \tilde{W}_i^{123} > 0$ in the concerned range $(a, \tilde{a})$. Hence, in the range $(a, \tilde{a})$ we have

$$\hat{W}_i^{123} - \tilde{W}_i^{123} > 0 \quad \forall \quad i = 1, 2.$$

Now, for Country-3 we have –

$$\hat{W}_3^{123} > \tilde{W}_3^{12} \text{ as } a_3 < \tilde{a}$$

and

$$\tilde{W}_3^{123} > \tilde{W}_3^{12} \text{ as } a_3 < \bar{\bar{a}}$$

Hence, as solved earlier using backward induction we arrive at $\mathbf{CU}_{\mathbf{123}}$ as the unique SPNE in this region.

Region IV: $\boldsymbol{\tilde{a} < a_3 < \bar{\bar{a}}}$

For the existing members (Country-1 and 2) we have the following welfare rankings

$$\hat{W}_i^{123} > \hat{W}_i^{12} \quad \forall \quad i = 1, 2 \text{ as } a_3 > \bar{a} \text{ and}$$

$$\tilde{W}_i^{123} > \hat{W}_i^{12} \quad \forall \quad i = 1, 2 \text{ as } a_3 > a$$

However, for Country-3 we get the following welfare rankings

$$\tilde{W}_3^{123} > \tilde{W}_3^{12} \text{ as } a_3 < \bar{\bar{a}} \text{ and}$$

$$\tilde{W}_3^{12} > \hat{W}_3^{123} \text{ as } a_3 > \tilde{a}$$

Hence, in this region we have $\tilde{W}_3^{123} > \tilde{W}_3^{12} > \hat{W}_3^{123}$. As a result, Country-3 will not accept the offer of joining CU$_{123}$ but will accept the offer of accession to FTA$_{123}$.

Again, from the welfare rankings of the initial existing member countries we can use backward induction method so as to reach the SPNE of the game and in this region **FTA$_{123}$** turns out to be the unique SPNE.

Region V: $a_3 > \bar{\bar{a}}$

As $a_3 > \bar{\bar{a}}$, Country-3 will not opt for both FTA or CU with existing FTA$_{12}$ member countries. Now, as we know that $\hat{W}_i^{12} > \tilde{W}_i^{12} (\forall\, i = 1, 2)$ irrespective of the market size of the bigger non-member Country-3, **CU$_{12}$** turns out to be unique SPNE of the game in this region.

References

Aghion P, Antras P, Helpman E (2007) Negotiating free trade. J Int Econ 73(1):1–30

Ahlström I, Stålros C (2005) Integration, trade pattern and intra-industry trade in ASEAN. Department of Economics, School of Economics and Management, Lund University

Alesina A, Angeloni I, Etro F (2005) International unions. Am Econ Rev 95(3):602–615

Berglof E, Burkart M, Friebel G, Paltseva E (2008) Widening and deepening: reforming the European union. Am Econ Rev Pap Proc 98(2):133–137

Brander JA (1981) Intra-industry trade in identical commodities. J Int Econ 11(1):1–14

Brander JA, Krugman P (1983) A 'Reciprocal Dumping' model of international trade. J Int Econ 15(3–4):313–321

Brulhart M (2009) An account of global intra-industry trade, 1962−2006. World Econ 32(3):401–459

Egger P, Larch M (2008) Interdependent preferential trade agreement memberships: an empirical analysis. J Int Econ 76(2):384–399

Freund C (2000) Different paths to free trade: the gains from regionalism. Q J Econ 115(4):1317–1341

Gaulier G, Lemoine F, Ünal D (2012) The rise of emerging economies in the EU15 trade. European J Comp Econ 9(1):133–175

Gil S, Llorca R, Martinez-Serrano JA (2008) Assessing the enlargement and deepening of the European union. World Econ 31(9):1253–1272

Gilligan MJ (2004) Is there a broader-deeper trade-off in international multilateral agreements? Int Org 58(3):459–484

Hausken K, Mattli W, Plumper T (2006) Widening versus deepening of international unions. http://papers.ssrn.com/Sol3/papers.cfm?abstract_id=953135
Hurley DT (2003) Horizontal and Vertical intra-industry trade: the case of Asean trade in manufactures. Int Econ J 17(4):1–14
Hu X, Ma Y (1999) International intra-industry trade of China. Weltwirtschaftliches Archiv 135 (1):82–101
IMF (2012) http://www.imf.org/external/pubs/ft/weo/2012/update/02/index.htm
Khalifah NA (1996) AFTA and intra-industry trade. ASEAN Econ Bull 12(3):351–368
Kemp M, Wan H (1976) An elementary proposition concerning the formation of customs unions. J Int Econ 6(1):95–107
Kennan J, Riezman R (1990) Optimal tariff equilibria with customs unions. Canadian J Econ 23 (1):70–83
Lee J-W, Park I, Shin K (2008) Proliferating RTAs: why and whither? World Econ 31(12): 1525–1557
Lorz O, Willmann G (2008) Enlargement versus deepening: the trade-off facing economic unions. CESifo working paper no. 2455
Menon J (1996) Intra-industry trade and the ASEAN free trade area. Pacific Economic Papers, No. 251, 1–20
Park I, Park S (2008) Free trade agreements versus customs unions: an examination of East Asia. Asian Econ Pap 8(2):119–139
Raff H (2001) Preferential trade liberalization under cournot competition. Rev Int Econ 9(3): 455–461
Seidmann DJ (2009) Preferential trading arrangements as strategic positioning. J Int Econ 79(1):143–159
WTO RTO database. http://rtais.wto.org/UI/PublicMaintainRTAHome.aspx
Yi S-S (1996) Open regionalism and world welfare. Eastern Econ J 22(4):467–475
Yi S-S (2000) Free-trade areas and welfare. Rev Int Econ 8(2):336–347

Deal Breaker or the Protector of Interests of Developing Countries? India's Negotiating Stance in WTO

Parthapratim Pal

Abstract The initial years of WTO promised that India would be a net gainer since the benefits accruing to them from the liberalisation of the three key sectors, namely, agriculture, textiles and services would more than offset the expected losses from removal of quantitative restrictions and imposition of a stricter intellectual property rights regime. However, the implementation experience of the WTO has been less than satisfactory for developing and least developed countries. So, when the Doha Development Round was launched, it was emphasised that the new round would take into account the development needs of poorer countries and would address the implementation issues of the Uruguay Round agreement. This chapter will analyse India's engagement in the Doha Round of trade talks in the light of its experience with the WTO regime. It will also look into the changing global economic landscape including the proliferation of the regional trade agreements and a regime of increasing commodity prices to analyse India's evolving negotiating position in the Doha Round.

1 Introduction

The World Trade Organization (WTO) celebrated its twentieth anniversary in 2015. On its website, WTO lists the achievements of these 20 years. The list is impressive. It says that during this 20-year period, WTO has helped to boost global trade growth, resolved numerous trade disputes and supported developing countries to integrate into the international trading system. It also highlights that WTO has played a crucial role during the global financial crisis of 2008. In a period of recession, there is a possibility that policymakers may get into a protectionist mindset. This has happened before with disastrous consequences for world trade.[1]

[1]During the Great Depression of 1930s, between 1929 and 1933 global trade shrank 65 % in dollar value and 25 % in unit volume.

P. Pal (✉)
Economics Group, Indian Institute of Management Calcutta, Kolkata, India
e-mail: parthapal@iimcal.ac.in

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_8

WTO claims that its rules, its dispute settlement mechanism and the forum for trade negotiation that it provides have acted as "a bulwark against protectionism" and has avoided a collapse of world trade. These are indeed credible achievements and WTO can rightly claim credits for these. The popularity of WTO can be further highlighted from the fact that it has increased its membership from 128 countries during the Uruguay Round to 161 countries in 2015. Among the new members, which acceded to WTO during this period, are big countries like China and Russia. Currently WTO members account for around 98 % of global trade. There are also more than 20 countries which are trying to get membership of WTO. Once the accession formalities of these observer countries are completed then WTO membership will cover almost the entire globe. These are major achievements and WTO should justly be proud of these.

There are some major pitfalls too. WTO has long been accused of showing bias to more developed countries. Concerns were raised about implementation problems of the agreements made under the Uruguay Round. This was acknowledged while launching the Doha Development Round in 2001. The official declaration of the Doha Round says:

> We attach the utmost importance to the implementation-related issues and concerns raised by members and are determined to find appropriate solutions to them. Para 12, Doha Ministerial Declaration.[2]

However, the Doha Round of trade talks has proved to be a major setback for WTO. It has been going on for about 15 years now and the end is still nowhere in sight. After the WTO talks collapsed in 2008, not much progress in negotiations have been made. There have been some agreements in a couple of areas but even after so many years, consensus in major areas of negotiations has eluded the WTO members. In a meeting with all WTO member countries on 1 June 2015, the director-general of WTO Roberto Azevedo says: "However, on the basis of the discussions I have had over recent weeks, I am becoming increasingly concerned that we are not making the progress that is needed in the key areas of agriculture, industrial products and services".[3] He further suggested that WTO members are lacking the required convergence in negotiations which may lead to a successful completion of the Doha Round of trade talks by December 2015.

As the Doha Round of trade talks are muddling through without any sign of completion, global interest in the multilateral trading system is on a decline. In the present era of data analytics, Google Trend has emerged as a useful tool to understand change of interest about a certain topic over a period of time. A Google Trend search using the word 'WTO' shows a rapidly declining interest about it (Fig. 1). Also, global trade is stuttering in the last few years. International trade statistics released by WTO in April 2015 indicate that growth in the volume of

---

[2]Available at: https://www.wto.org/english/thewto_e/minist_e/min01_e/mindecl_e.htm.

[3]Available at: https://www.wto.org/english/news_e/news15_e/hod_01jun15_e.htm.

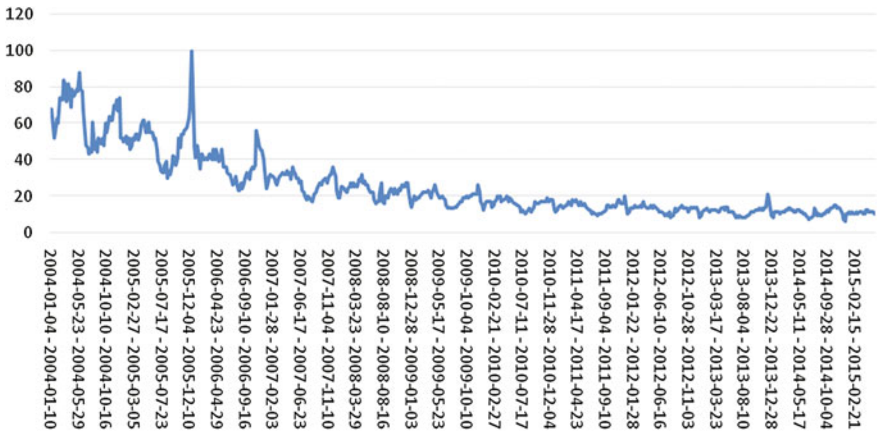


Fig. 1 Result from Google trend of relative interest in the search term 'WTO'. *Source* Google trends

world merchandise trade will be around 3.3 % in 2015 and will remain well below the annual average of 5.1 % posted since 1990. The modest trade growth in 2014 marks the third consecutive year in which trade grew less than 3 %. Trade growth averaged just 2.4 % between 2012 and 2014. It is likely that global demand is still recovering from the recession and slow recoveries of some major countries have affected the global trade growth. However, some amount of responsibility can be attributed to the lack of progress of the Doha Round and the inability of WTO to take trade liberalisation forward in the last 15 years.

India has played a much more active role in the Doha Round of negotiations than it did in any previous rounds of trade talks. Right from the inception of this round, India has emerged as a key player. Tough negotiating stance adopted by India in certain issues has contributed in stalling the Doha Round of trade talks a few times. Consequently, India has often been accused of brinksmanship and been criticised for jeopardising the entire Doha Round. India, on the other hand, has been almost prided itself in stopping bad and iniquitous trade agreements.[4] This chapter will analyse India's engagement in the Doha Round of trade talks in the broader perspective of its experience with the WTO regime. It will also look into the changing global economic landscape, including the proliferation of the regional trade agreements and a regime of increasing commodity prices to analyse India's evolving negotiating position in the Doha Round.

[4]For example, see 'Doha: India accuses US of sacrificing world's poor at trade talks', The Guardian, July 31, 2008. Available at: http://www.theguardian.com/world/2008/jul/31/wto.india.

2 India in GATT and WTO

Though India pursued inward-looking economic policies during most of its planning period, it was a founder member when the General Agreement on Tariffs and Trade (GATT) was established in 1947. India continued with inward-looking policies until about the early 1980s when it started to open up. The major change occurred in 1991 when India had to approach the International Monetary Fund (IMF) for structural adjustment loans to tide over a balance of payments crisis. As conditions for the loan, IMF imposed several measures to make the Indian economy more open and liberalised. This ushered in a period of rapid trade liberalisation for India. India's drive towards a more open trade regime also coincided with the Uruguay Round of trade talks and the establishment of the WTO in 1994. As a member of GATT, India also became a founder member of the WTO.

During the mid-1990s, policymakers widely believed that the WTO would be a vast improvement over GATT and the establishment of the new multilateral trading system would help India take advantage of the growing global demand. There was a perception that India would be a net gainer from the WTO regime since the benefits accruing to them from the liberalisation of the three key sectors, namely, agriculture, textiles and services, would more than offset the expected losses from India's commitments to reduce trade barriers and adopt a more stringent regime of Intellectual Property Rights (IPRs). Indian policymakers expected that increased market access in these three key sectors would allow the country to follow an open and export-led development policy. Since an overwhelming proportion of workers in India are involved in agriculture, textiles and services, it was also expected to significantly boost employment. However, the implementation experience of the WTO during the period 1995–2000 has been less than satisfactory for developing and least developed countries. For example, in agriculture, developing countries were disappointed that the Agreement on Agriculture (AoA) did not lead to any meaningful reduction in subsidies. In textiles, the major liberalisation was back-loaded until the Multi Fibre Agreement expired in 2005. In services, countries like India felt that they did not get enough market access for their people, even professionals. So, when the Doha Development Round was launched in 2001, it was emphasised that the new round would take into account the development needs of poorer countries and would address the implementation issues of the Uruguay Round agreement.

In the Doha Round of trade talks, India has been actively involved in negotiations in all key areas of the WTO. One of the defining features of the Doha Round of trade talks is that the level of involvement of member countries in trade negotiations has been higher than ever before. In the past, most negotiations happened among the so-called QUAD (USA, European Commission, Japan and Canada) countries. The only other significant pressure group was the Cairns Group.[5] Most developing and least developed countries had much lower level of participation in the trade talks. However, in the Doha Round, developing countries have participated in the negotiations both as individual countries and also as groups of countries with similar negotiating positions in different areas of trade negotiations. Formation of such groups has given countries the 'pack power' and increased bargaining power of smaller countries in the Doha Round. India is part of various such negotiating groups. A look at India's engagement in different groups shows that India is involved with three different country groupings in agriculture negotiations. These groups are Asian Developing Members (31 Members),[6] G20 (23 Members)[7] and G33 (46 Members).[8] In Non-Agricultural Market Access, it is involved with Asian Developing Members and NAMA11 (10 Members).[9] In discussions regarding anti-dumping; subsidies and countervailing measures, including fisheries subsidies; and regional trade agreements (this set of issues are termed as 'Rules' in shorthand at the WTO negotiations) and TRIPS, India is involved with the group Asian Developing Members (Table 1).

The next few sections will discuss India's position in this round of trade talks in detail.

---

[5]The Cairns Group is a group of countries which negotiated for a more open agricultural trade regime since the Uruguay Round of trade talks. Members of the Cairns Group are: Argentina, Australia, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Guatemala, Indonesia, Malaysia, New Zealand, Pakistan, Paraguay, Peru, the Philippines, South Africa, Thailand and Uruguay. Hungary (now part of the European Community) and Fiji were founding members of the Cairns Group, but have since withdrawn.

[6]Members of this group are: Bahrain, Kingdom of, Bangladesh, Brunei Darussalam, Cambodia, China, Chinese Taipei, Hong Kong, China, India, Indonesia, Jordan, Korea, Republic of, Kuwait, the State of, Kyrgyz Republic, Lao People's Democratic Republic, Macao, China, Malaysia, Maldives, Mongolia, Myanmar, Nepal, Oman, Pakistan, Philippines, Qatar, Saudi Arabia, Kingdom of, Singapore, Sri Lanka, Thailand, Turkey, United Arab Emirates, Viet Nam.

[7]G-20 is a coalition of developing countries in WTO that are negotiating for ambitious reforms of agriculture in developed countries with some flexibility for developing countries. This group is not to be confused with the G-20 group of finance ministers and central bank governors. Members of this group are: Argentina, Bolivia, Plurinational State of, Brazil, Chile, China, Cuba, Ecuador, Egypt, Guatemala, India, Indonesia, Mexico, Nigeria, Pakistan, Paraguay, Peru, Philippines, South Africa, Tanzania, Thailand, Uruguay, Venezuela, Bolivarian Republic of, Zimbabwe.

[8]Members of this group are: Antigua and Barbuda, Barbados, Belize, Benin, Bolivia, Plurinational State of, Botswana, Côte d'Ivoire, China, Congo, Cuba, Dominica, Dominican Republic, El Salvador, Grenada, Guatemala, Guyana, Haiti, Honduras, India, Indonesia, Jamaica, Kenya, Korea, Republic of, Madagascar, Mauritius, Mongolia, Mozambique, Nicaragua, Nigeria, Pakistan, Panama, Peru, Philippines, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Senegal, Sri Lanka, Suriname, Tanzania, Trinidad and Tobago, Turkey, Uganda, Venezuela, Bolivarian Republic of, Zambia, Zimbabwe.

[9]Members are: Argentina, Brazil, Egypt, India, Indonesia, Namibia, Philippines, South Africa, Tunisia, Venezuela, Bolivarian Republic of.

Table 1 Various negotiating groups in the Doha round where India is a member. *Source* https://www.wto.org/english/tratop_e/dda_e/negotiating_groups_e.htm

Negotiating area	Group name
Agriculture	Asian developing members, G20, G33
Non-Agricultural market access	Asian developing members, NAMA 11
TRIPS	Asian developing members
Rules	Asian developing members

2.1 Agreement on Agriculture and India

When the WTO deal was signed, the Agreement on Agriculture (AoA) was seen as a major breakthrough in multilateral trade negotiations. Early projections of the likely effect of the AoA on world markets predicted that a reduction in domestic support and export subsidies in the developed countries would lead to a deepening of world trade in agriculture, a spatial redistribution of agricultural production, an increase in the share of developing countries in global agricultural exports and greater transparency in agricultural trade. However, closer examination suggested that there would be no major reduction in subsidies in the OECD countries. The provisions in the agreement allowed countries to comply with WTO rules but continue with their farm subsidies. Consequently, the AoA produced modest results. Developed countries continued with their domestic and export support measures, while developing countries continued to have high tariff barriers for this sector. There was an in-built provision in the WTO AoA (under Article 20 of the AoA) that committed members to begin negotiations on continuing the reform at the end of 1999 (or the beginning of 2000). So, when this mandated negotiation on agriculture started in 2000, India was expected to adopt a position in favour of greater liberalisation in agriculture. However, India was ambivalent in its approach. Although India belongs to the G-20 group of developing countries in the WTO that pushes for major liberalisation in agriculture, including ambitious reductions in market access barriers and farm subsidies, India is also a member of the G-33 group, which argues for greater flexibility for developing countries to undertake limited market opening in agriculture.

In its proposal to the WTO (WTO document number G/AG/NG/W/102) in January 2001, India emphasised that the negotiations on agriculture should focus on both the trade and non-trade concerns of developing countries. Along with trade issues, such as domestic support, export subsidies and market access; non-trade concerns, such as food and livelihood security, should also be addressed. India also pointed out that there are a large number of low income, resource-poor agricultural producers in the country, and India needs to guard their interests. These concerns have guided India's approach to the Article 20 re-negotiations on agriculture and subsequently to the agriculture negotiations in the Doha Development Round.

Due to this defensive stance, India has been guarded in its approach towards reducing agricultural tariffs. India's defensive stance may appear counter-intuitive because India has unilaterally liberalised its agriculture trade regime much beyond its WTO commitments. In India, applied tariff rates in agricultural products are much lower than the bound rates. For agricultural products, the average MFN applied tariff rate is around 35 % and it ranges between 0 and 150 %. India's bound rates in agricultural products range from 0 to 300 % and there is significant gap (or water content) between the applied and bound rates for most products. India has brought down tariffs significantly in edible oil and industrial raw materials.[10] Despite this, India has been reluctant to aggressively pursue tariff liberalisation in agriculture. The Ministry of Commerce website mentions that India will not be able to undertake overall tariff reductions on bound rates of more than 36 %. India is also in favour of several additional flexibilities, such as self-designation for a number of Special Products,[11] greater flexibility on Sensitive Products and simpler use of Special Safeguard Mechanisms (SSMs).

Regarding market access initiatives by other countries, Indian exporters have repeatedly pointed out that although the average level of tariff is low in developed countries, there are still tariff peaks and tariff escalations. Moreover, there are major non-tariff barriers in terms of food standards and technical barriers to trade. Some of these issues have been highlighted by the Ministerial declaration of the G-20 group (WTO document number WT/MIN(11)/19 dated 16 December 2011) at the Geneva Ministerial Meet.

At the same time, India calls for substantial reduction in domestic support measures in the agriculture sector of developed countries, with tighter disciplines on product specific limits on AMS and the Blue Box.[12] It also wants the complete abolition of export subsidies by developed countries and a phase out of export subsidies by developing countries.

India's negotiating stance in agriculture has been criticised by researchers such as Hoda and Gulati (2007), who argue that developing countries, including India, should be less preoccupied with S&D (special and differential) treatment[13] in agriculture. Rather, these countries should negotiate more aggressively to eliminate

---

[10]However, in 1999 India increased its bound rates through Article XXVIII tariff negotiations on some agricultural commodities, mostly cereals, including rice and wheat.

[11]These are agricultural products that have implications for food security, livelihood security and rural development. These products will attract a lower tariff cut than other products. This is an S&D measure since only developing countries can designate Special Products.

[12]WTO divides agricultural subsidies into three categories. Domestic subsidies for agriculture that are considered to distort trade are called Amber Box subsidies. These subsidies are subject to reduction commitments. Amber Box is calculated as Aggregate Measurement of Support (AMS). Farm subsidies that are decoupled from production or are not production-enhancing are called Blue Box subsidies. Currently, there is no limit on Blue Box subsidies. Green Box subsidies are domestic support for agriculture that do not distort trade or at most cause minimal distortion. These subsidies are allowed without limit.

[13]Special and Differential Treatment refers to provisions within the WTO that give developing and least developed countries (LDCs) special rights and allow developed countries to treat developing countries and LDCs more favourably than other WTO members. These special provisions include, for example, longer periods for implementing agreements and commitments, or measures to increase developing countries' trading opportunities.

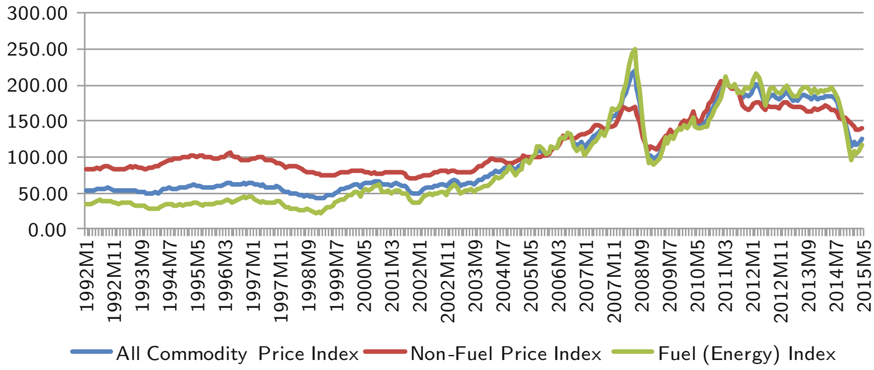


Fig. 2 International monetary fund commodity price indices, (Base 2005 = 100). *Source* International Monetary Fund

domestic and export support in developed countries. By being defensive in their market access commitments, countries like India reduce the overall level of ambition for the negotiations on agriculture.

It should also be pointed out that special products, sensitive products and SSMs are instruments that are likely to help a country during a regime of declining prices. Since 2003–04, international commodity prices have been mostly rising. Though commodity prices have shown some tendency of decline since August 2014, the decline has been much sharper for fuel products than for non-fuel products (see Fig. 2). The figure also shows that since 2008, commodity prices have shown very high volatility. However, studies tend to indicate that in spite of these volatilities, the world may have moved into a period of high international commodity prices.[14] If this hypothesis is correct then it implies that in this new price regime the above-mentioned instruments will be less useful for price stabilisation. A high price regime may also mean lower resistance at the international level against domestic and export support given by developed countries. On the flip side, a high price regime also reduces the need for domestic and export subsidies. It is not surprising that in the WTO negotiations, agriculture is no longer considered the main hurdle. Negotiations are now stalled by fundamental differences in non-agricultural market access.

Another issue that may influence India's outlook towards agriculture is that the country has faced relatively high food price inflation in the past few years due to domestic demand–supply imbalances and high international commodity prices. For a poor country like India, food price inflation is politically difficult to handle, and if the domestic demand–supply imbalance continues, it may have an impact on India's approach towards agricultural imports. Some amount of this concern was felt in Bali Ministerial Meet of WTO when India almost single handedly blocked an

[14]For example, see Hermann (2009), FAO (2011).

accord on trade facilitation as it felt that trade facilitation could not be de-linked from other aspects of the Doha Round. The position taken by India arose from strategic concerns in relation to another accord on food security and public stockholding in agriculture tabled during the same meeting. Indian government perceived it as a possible impediment for ensuring food and livelihood security in the country. Subsequently, in November 2014, an agreement was reached between USA and India which allowed the trade facilitation accord to be implemented.

2.2 Non-Agricultural Market Access and India

When the Doha Round started, it was believed that negotiations on Non-Agricultural Market Access (NAMA) would be less complicated than in the other sectors for two reasons. First, industrial tariffs are much lower across the world than tariffs on agricultural goods and, second, the Doha Round of trade talks was the ninth round of trade negotiations on industrial goods and it was believed that most of the major issues had been solved.

During the NAMA negotiations, it was decided that a Swiss formula-based tariff reduction strategy would be adopted.[15] The coefficient of the Swiss formula is critical in arriving at a reduction commitment and it was decided in WTO trade talks that the Doha Development Round would take into account the special needs and interests of developing countries and allow less than full reciprocity in reduction commitments. Consequently, it was decided that there would be different coefficients for developed and developing countries. During the negotiations, there was a convergence of opinion that the coefficient would have a value of 8 for developed countries, whereas for developing countries, there would be a three-tiered coefficient with values of 20, 22 and 25 to address flexibilities for these countries. India had agreed to this formula. The Hong Kong Ministerial declaration recognised that members were pursuing sectoral tariff reduction initiatives to supplement the formula-based approach, but emphasised that participation in such sectorals should be voluntary.

However, the global economic crisis and continued trade deficit of the US have given rise to strong mercantilist sentiments in some countries. Its manifestation can be seen in the currency war between China and the US, the increasing protectionist

[15]A Swiss Formula for tariff reductions is given by the equation:

$$Z = CX/(C + X)$$

where

X initial tariff rate

C coefficient

Z resulting lower tariff rate (end of period).

measures adopted by developed countries since the financial crisis and in the sudden hardening of country positions in WTO trade talks on industrial goods. The US and other developed countries now argue that large developing countries, such as Brazil, China and India, should participate in sectoral tariff reductions on a mandatory basis. The sectors picked by the US were chemicals, electrical and electronics and industrial machinery. Later, more sectors were added and a WTO document on NAMA sectoral negotiations (WTO document number TN/C/14 dated April 21, 2011) indicate that currently seven sectors are earmarked: chemicals, industrial machinery, electronics and electrical products, enhanced healthcare, forest products, raw materials and gems and jewellery. This document also highlights that the sectoral initiatives are driven by developed countries, namely, Australia, the European Union (EU), Japan and the US. The three developing countries (Brazil, China and India) indicated that they are not seeking further market access through sectorals and they want the Swiss formula to be the main determinant of the overall level of ambition in the NAMA negotiations. Sectorals should be seen as a supplement to the tariff cuts achieved through the formula, and participation in sectorals should be non-mandatory.

Another development in the NAMA negotiations goes against countries like India, which have unilaterally liberalised their tariff rates. The GATT and WTO rules traditionally treat the bound rate as the base rate for future tariff cuts. However, in the Doha Round, there has been a demand from developed countries to reduce tariff rates from the present applied rates. This is discriminates against a country like India that has unilaterally reduced its applied tariff rates much below its WTO commitments.

Given the present state of the international economy, it is unlikely that India will accept these demands in the NAMA negotiations. These new demands go against the basic premise of the Doha Development Round because major developed countries are moving away from the concept of less-than-full reciprocity by developing countries, and are pushing "emerging countries to "catch up" with developed members regarding the level of market opening" (p. 2).[16]

The issue of non-tariff measures (NTMs) is also important for India. Increasingly, non-tariff measures are being used as protectionist tools in several developed countries. The use of safeguard measures, such as Sanitary and Phytosanitary (SPS) measures, Technical Barriers to Trade (TBT) and anti-dumping (AD),[17] are used to discriminate against imports from developing countries like India. In addition, new forms of NTMs, such as maintaining undervalued currency, certification and documentation, are being used. To make international trade more transparent, India seeks better rationalisation of these measures.

[16]Report by the Director-General on his Consultations on NAMA Sectoral Negotiations, WTO document number TN/C/14 dated April 21, 2011.

[17]India's position against the misuse of AD is somewhat weak since India is a major user of AD measures.

2.3 Trade in Services and India

Services are a very important sector for India. In 2009–10, the services sector accounted for over 50 % of its GDP. The sector grew at an average annual rate of 10 % in 2009–10 and contributed nearly 60 % to the overall growth of the economy. The leading subsectors are financial services, trade and transport and communications. Trade in services is also growing rapidly. India's services exports increased from US$ 25 billion in 2003–04 to US$ 90 billion in 2007–08. According to the World Trade Report 2014, India exported US$ 150 billion worth of services in 2014.

In the WTO, India is a major proponent of services liberalisation in the Doha Round. India's negotiating position reflects its core interest in getting market access and non-discriminatory treatment from its trading partners, especially developed country trading partners, in sectors, such as computers and related services, and in modes of trade, such as Mode 4 (temporary movement of people) and Mode 1 (cross-border trade) where the country has an export interest. To prove its offensive interests in services liberalisation, India offered to undertake commitments in several new sectors, such as distribution services, education, environmental services, transport services (maritime and air) and professional services (architectural services, integrated engineering services, medical and dental services), and offered to improve its Uruguay Round commitments in sectors such as computer and related services, R&D services and in some professional services (accounting, auditing and bookkeeping services, engineering services, etc.) in its Revised Offer submitted to the WTO in 2005. This was one of the best revised offers submitted to the WTO.

In its requests to trading partners in the Doha Round, India is looking for broad-based commitments in Mode 1 from a critical mass of countries in some key areas. The objective is to freeze the existing levels of liberalisation in those countries to pre-empt potential future hurdles in outsourcing business to India. Although India has a strong export interest in Mode 4, in the Doha Round developed country members have taken limited commitments in this Mode. Moreover, these commitments are restricted by various measures, such as economic needs tests, non-recognition of qualifications and visa restrictions. It is of the utmost importance to India that in the Doha Round there is progress on the development of discipline in domestic regulations involving qualifications and licencing requirements and procedures, without which Mode 4 access gets severely restricted.

Along with the traditional request-offer method used in GATS, the Doha Round of trade talks also uses a plurilateral negotiation approach to further liberalise trade in services. So far, 22 plurilateral groups have been formed at the WTO in service sectors/modes. India is the coordinator of the plurilateral requests on Mode 1 (cross-border supply) and Mode 4 (Movement of Natural Persons). India is a cosponsor of plurilateral requests in computer and related services (CRS).

However, the quality of offers in services in the Doha Round has been disappointing. Borchert et al. (2011) pointed out that several of the offers made in the Doha Round are actually more restrictive than the present policies. They conclude "the best offers submitted so far as part of the Doha negotiations improve on Uruguay Round commitments by about 10 % but are still, on average, twice as restrictive as actual policies. At present, Doha offers not greater access to markets but a weak assurance that access will not get worse". Similar observations have been made in the April 2011 report by the Chairman to the services trade negotiation committee (WTO document Number TN/S/36 dated April 21, 2011). The report highlights the concern of countries such as India that no progress has been made in the market access negotiations for services since the July 2008 Signalling Conference. The gaps between offers or signals and bilateral and plurilateral requests or applied regimes were still substantial. Moreover, there was a lack of clear signs of market openings in sectors and modes of interest to developing countries, particularly in Modes 1 and 4.

India believes that to take the negotiations forward, a fresh round of offers need to be tabled at the WTO by member countries. An ambitious proposal in services should be an essential part of the Doha agreement. However, India has opposed some recent plurilateral negotiations on trade in services. Recently, there have been talks about an International Services Agreement (ISA) by a group called Really Good Friends of Services (RGFS) that is pushing for ambitious liberalisation in trade in services. This group includes Australia, Canada, Colombia, Costa Rica, the EU, Hong Kong China, Israel, Japan, Mexico, New Zealand, Norway, Pakistan, Peru, South Korea, Switzerland, Separate Customs Territory of Taiwan, Penghu, Kinmen and Matsu, Turkey, and the US. So far, large emerging economies such as India, Brazil and South Africa have stayed away from these plurilateral negotiations. In a meeting of leaders from Brazil, Russia, India, China, and South Africa in New Delhi in March 2012, the leaders of these countries warned: "We do not support plurilateral initiatives that go against the fundamental principles of transparency, inclusiveness, and multilateralism… We believe that such initiatives not only distract members from striving for a collective outcome but also fail to address the development deficit inherited from previous negotiating rounds" (Bridges Weekly, ICTSD, 2012). However, Hoda (2012) points out that the approach adopted by the RGFS may not be GATS-inconsistent as there are provisions in Article V of GATS to form such an agreement, but he agrees that such an agreement may further weaken the Doha Round of talks. To alleviate this problem, he suggests that if these countries manage to secure an agreement among a critical mass of members and then extend the benefits on an MFN basis to all WTO members, the effort may help the multilateral negotiations. He argues that instead of staying out of this approach, India should get involved and ensure all modes of supply, particularly the movement of natural persons, are covered in this possible agreement.

3 Which Way Is the Doha Round Heading?

Overall, the Doha Round of trade talks is not progressing well. The negotiations have been repeatedly stalled on issues related mainly to agriculture subsidies and non-agricultural market access. It has been 14 years since the Doha Declaration and still the negotiations show no sign of completion. However, it is not very surprising that the Doha Round is taking so long to get finalised. In the earlier rounds of multilateral trade talks, negotiations were done by a few developed countries and most developing countries had little or no role in these negotiations. In the Doha Round, the number of active players in negotiations has increased significantly. Rodrik (2008) points out that in the Doha round there are 153 countries (161 countries now), of which around 60 or 70 are actively involved in the negotiations. As the WTO works on a 'consensus' approach, every member has a potential veto power.[18] Therefore, it is not surprising that this round of trade negotiations is dragging on.

Moreover, WTO is facing challenges from a number of new developments in international business and trade. As the Doha Round started in 2001, its basic tenets were based on the pattern of international trade of 1990s. But since then, rapid changes in business and global dynamics have posed a number of new challenges for WTO. Some of major changes faced by WTO are

1. **Rise of China**: China was not a member when WTO was formed in 1995. It became a member only in 2001. But in spite of that, China's growth through the last two decades has been amazing. In 1990, China had a share of around 1.8 % in global merchandise exports which by 1995, climbed up to 2.88 %. While acceding to WTO in 2001, China's share in global exports reached 4.3 %. Since then China has managed to improve its market share at a very fast pace and in 2014 it had 11.74 % share in international merchandise exports. It is presently the biggest exporter of merchandise goods in the world. China has very strong presence in trade in commercial service as well. It has 4.41 % market share in global commercial services exports and is ranked fourth among commercial services exporters among the world. The rise of China as a global economic superpower was largely driven by its growing dominance in international trade. This meteoric rise of China in the last two decades has altered patterns of global trade and power balances quite significantly. Along with China, some other developing countries like Brazil, India and South Korea have also made their marks felt in international economic area. Emergence and assertiveness of new players have changed the dynamics of global trade negotiations quite significantly.
2. **Recession and Protectionism**: A major shock to the global economy came in 2008–09, when the subprime crisis and financial contagion across the globe led to the biggest global recession since the great depression of the 1930s. Not surprisingly, international trade also observed a massive contraction. In 2009,

[18]Although the WTO has the provision of a 'one-country one vote' system in its constitution, so far voting has not been used in the WTO.

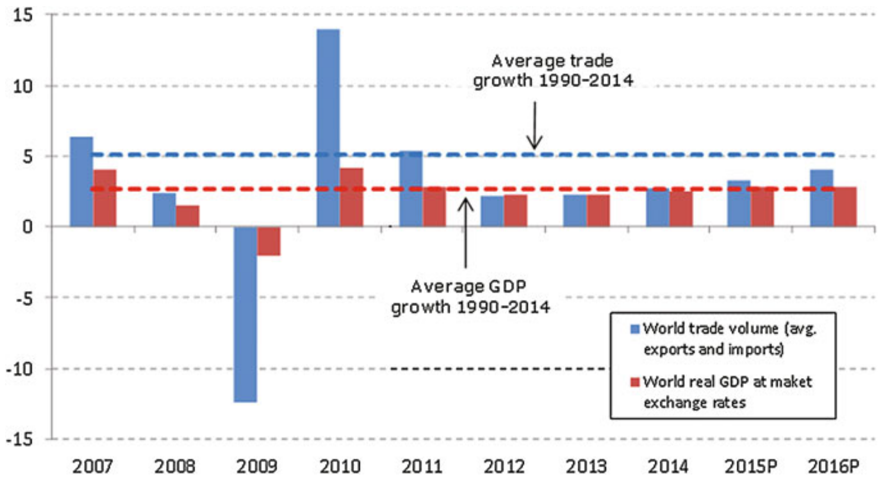


Fig. 3 Growth in volume of world merchandise trade and real GDP, 2007–16 (Annual % change). *Source* WTO

global merchandise trade declined sharply from US$ 16.1 trillion to US$ 12.4 trillion, registering a negative growth rate of more than 22 % (Fig. 3). Trade in commercial services also declined by about 11 % during the same year. Such a decline in the value of international trade has not happened for a very long period.

This massive recession and the contraction of world trade led to a sharp increase in protectionist sentiments among the developed countries. Studies done by WTO reported a steady rise in imposition of trade-restrictive measures in developed countries. This led to the possibility of a global breakout of protectionism like it happened in the 1930s during the Great Depression. However, presence of a multilateral trade monitoring body like the WTO and concerted effort by the G-20 countries prevented such a scenario. However, the recovery of international trade has not been smooth. A number of factors, including the Eurozone crisis, have slowed down exports growth since 2012 (Fig. 3).

3. **Massive rise in trade in intermediate goods and emergence of value chains**: A defining feature of modern international trade is the growing integration of countries in global value chains. In global value chains, companies divide their operations across the world depending upon the comparative advantage of each location. As a result, it relocates the processes starting from the design of the product and manufacturing of components to assembly and marketing in different countries across the globe. This leads to international production chains. For example, over the years, China has emerges as the final assembly centre from global manufacturing. A study by WTO and IDE-JETRO (WTO IDE-JETRO 2011) shows that about 75 % of China's imports are intermediate goods but its export of intermediate goods is one of the lowest among Asian countries.

This trade pattern is called by Baldwin (2011a, b) as the '21st Century trade' which is characterised by the 'trade investment-services' nexus. According to Baldwin,

> the nexus reflects the intertwining of: i) trade in parts and components, ii) international movement of investment in production facilities, training, technology, and long-term business relationships, and iii) demand for services to coordinate the dispersed production, especially infrastructure services such as telecoms, internet, express parcel delivery, air cargo, trade-related finance, customs clearance. (Baldwin 2011a)

Baldwin further suggests that this '21st century trade' will require policy packages which are very different from traditional trade where countries generally trade in finished goods. For example, issues like trade facilitation and trade costs become more important for a country which is engaged in trade in intermediate goods and has relocated its production bases around the world. Similarly, stronger trade-investment nexus imply that issues like competition policy, investment treaties, IPR issues and restrictions on movement of capital may become more important than traditional trade policy issues like tariffs and NTMs.

WTO recognises the growth of global value chains through it's 'Made in the World' initiative but it finds itself lacking in tackling this type of international trade. As an organisation, WTO's basic tenets are still rooted in GATT—an informal organisation designed to monitor international trade of the 1940s. Therefore, its primary focus is still on instruments like quotas, tariffs and subsidies. Moreover, as a global behemoth, which runs on the basis of consensus, flexibility is not a key strength of WTO. Consequently, it is finding it difficult to keep pace with modern business and often finds itself trying to manage tomorrow's business and economy using yesterday's policy instruments. When global business and finance is changing ever so rapidly with technology and innovations, WTO may need to become much more flexible and adaptable if it wants to remain relevant. As even more modern technologies like 3-D printing emerge, new possibilities and policy challenges will have to be addressed by WTO.

4. **Commodity Prices behaved differently since 2003**: After a prolonged period of relative stagnation, international commodity prices are on an upswing since 2003–04. There are alternative hypothesis about what is pushing up commodity prices. One group of economists strongly believe that demand–supply imbalance is causing this upswing whereas others believe that commodities have emerged as an alternative to financial assets and speculation is the major reason for this upward movement of commodities. Expansionary monetary policies in developed countries are also said to be helping commodity prices.

Most of the WTO policies and safeguards like 'Special Products' and 'Special Safeguard mechanisms', are essentially geared to protect countries from the negative effects of declining commodity prices. In the present WTO texts, there are no substantive policy measures available to developing countries to protect them from increasing food prices. In other words, WTO lacks policies towards general price stabilisation. Its policies and tools are from an era where international commodity prices were going through secular decline. Those are much less relevant in the present scenario.

5. **Rise of Regionalism**: As the multilateral trade talks are not progressing, most countries are opting to increase trade through Regional Trade Agreements (RTAs). RTAs are defined as groupings of countries which are formed with the objective of reducing barriers to trade between member countries. Contrary to what the name suggests, these groupings or unions may be concluded between countries not necessarily belonging to the same geographical region. Regional agreements allow groups of countries to negotiate rules and commitments that go beyond WTO rules. For example, these agreements can include stricter TRIPS regulations, Bilateral Investment Treaties (BITs), wide range of and extraneous clauses like labour and environmental rules.

Regional trade agreements represent an important exception to the WTO's principle of non-discrimination (the Most Favoured Nation clause). According to the WTO rules, countries within an RTA (or, in other words, members of an RTA) can trade among themselves using preferential tariffs and easier market access conditions than what is applicable to other WTO Member countries. As a result, WTO Member countries that are not a part of the RTA are discriminated in these markets. Also trading within the regional trade blocks does not come under the purview of WTO. Therefore, WTO dispute settlement mechanisms are also not available to countries that form RTAs. In spite of all these drawbacks, WTO allowed formation of RTAs as it believed that such trade agreements may allow countries to liberalise and expand trade by choosing their own trade partners and adopting the right pace of trade liberalisation among its chosen trade partners. However, as trade among regional trade blocks are outside the ambit of WTO, proliferation of RTAs can reduce the policy relevance of WTO quite significantly. In the last few years, there have been talks about formation of mega trade blocks involving large number of countries. Examples of such mega trade blocks are The Regional Comprehensive Economic Partnership (RCEP)[19] and the Trans-Pacific Partnership (TPP).[20] Once these trade blocks become operational, WTO may struggle to remain relevant for many countries.

The confluence of these factors as well as the fact that Doha Round has not worked out well imply that there will be difficult days ahead for WTO. Moreover, proliferation of RTAs ensured that most WTO Members are party to multiple

[19]Regional Comprehensive Economic Partnership (RCEP) is a proposed free trade agreement (FTA) between the ten member states of the Association of Southeast Asian Nations (ASEAN) (Brunei, Burma (Myanmar), Cambodia, Indonesia, Laos, Malaysia, the Philippines, Singapore, Thailand, Vietnam) and the six states with which ASEAN has existing FTAs (Australia, China, India, Japan, South Korea and New Zealand).

[20]The Trans-Pacific Partnership (TPP) is a proposed regional regulatory and investment treaty. As of 2014, twelve countries throughout the Asia Pacific region have participated in negotiations on the TPP: Australia, Brunei, Canada, Chile, Japan, Malaysia, Mexico, New Zealand, Peru, Singapore, the United States, and Vietnam.

RTAs. WTO database on RTAs suggest that some WTO Members are participating in twenty or more RTAs. This has resulted in overlapping trade agreements. As a consequence, global trade is presently being governed by a spaghetti bowl of bilateral and multilateral trade agreements. Baldwin et al. (2007) point out that WTO has largely been an "innocent bystander" in this rapid rise of regionalism which potentially can erode the benefits of multilateralism by introducing more opacity in the system as well as by giving rise to a lopsided global trading system dominated by large trading powers.

4 India's Role in Doha Round: Running with the Hare and Hunting with the Hounds?

As discussed before, India has played a more proactive role in this round of WTO negotiations. It has held its ground against pressure and has so far followed the dictum that 'no deal is better than a bad deal'. But this tough stance has also invited criticism from the international community and media. This is best illustrated by the criticism India faced after the veto to the trade facilitation accord during the Bali Ministerial. Although domestically the move by the government was seen as a positive one, internationally India was seen as a deal breaker in WTO. In some cases it was also alleged that India's action was a setback against free trade.[21] While it has to be agreed that the veto was not a popular decision as only three countries, (Cuba, Venezuela and Bolivia) voiced support for India, India's concerns were not without merit. But the global media blitz against India made the country look somewhat isolated from other developing countries. While it is true that in that particular incident India did not have too many supporters, it is also true that in a number of other areas of negotiations in WTO, especially in areas like agricultural and NAMA, there is considerable solidarity and convergence among developing countries.

However, there is an unnecessary burden which India seems to have taken on itself is as the leader of developing countries. After the Bali Ministerial, the erstwhile Minister of Commerce and Industry said:

> I would like to conclude by saying that the Bali Ministerial meeting was a landmark one in the history of WTO. It re-affirmed India's leadership role amongst the developing countries and also demonstrated our diplomatic ability to build consensus. We were able to arrive at a balanced outcome which secures our supreme national interest. India was key to arriving at a breakthrough and shaping the first agreement since the creation of the WTO 18 years ago. India's constructive approach in negotiations was acknowledged by all member states. We have managed to retain the centrality of the development dimension in the Doha Round.[22]

In a visit to India during January 2015, the WTO director-general played along saying:

---

[21]'A Setback for Free Trade: Strangled at Birth', The Economist, August 1, 2014; 'World Trade: The Indian Problem', The Economist, November 21, 2013.

[22]Available at: http://pib.nic.in/newsite/PrintRelease.aspx?relid=101827.

> Countries look to India to raise issues of importance for developing countries. That role carries real responsibility. And I think we are seeing India take its rightful place at the centre of the world stage.[23]

This is not an easy burden to bear. Developing countries are not a homogenous group and there are considerable differences in negotiating interest among developing countries. Leading such a diverse group will be extremely difficult for India. Moreover, in spite of its population and GDP, India is a relatively small player in international trade. India's share in global exports was around 1.6 % in 2013. Its share in global services exports is around 3.3 %. There are many smaller economies with much larger trade volumes than India. In fact, China's annual merchandise exports is higher than India's GDP. In 2013 China's merchandise exports was US$ 2.2 trillion while India's GDP in the same year was US$ 1.87 trillion. This raises the doubt whether India is really big enough to carry the 'leadership' mantle of developing countries in international trade forums.

Finally, for countries like India, bilateral engagements can create complications for multilateral negotiations. Recent reports coming out of WTO seem to indicate that presently in WTO discussions on important issues are taking place behind closed doors in meetings among the so-called G5-Plus countries with the director-general of WTO as the chair. These countries are: US, the EU, China, India, Brazil, Australia and Japan. This is part of the 'Green Room meeting' convention in WTO where more interested and important parties have semi-formal negotiations among themselves. However, what is important here is to note that India also engages bilaterally with most of these countries of the core group. India has a comprehensive economic partnership agreement with Japan and negotiations with EU for a comprehensive trade and investment agreement is underway. India is also negotiating RTAs with Australia and New Zealand. Possibility of an Indo-US FTA is also around. As these bilateral agreements generally go much beyond WTO liberalisation requirements, it will be difficult for India to maintain a consistent negotiating ground between bilateralism and multilateralism, especially given the additional leadership responsibility that willy-nilly India has bestowed upon itself. India's strategy of running with the hare and hunting with the hounds in international trade forums will not be easy to sustain.

References

Baldwin R (2011a) 21st century regionalism: filling the gap between 21st century trade and 20th century trade rules. CEPR Policy Insight No. 56

Baldwin R (2011b) Global manufacturing value chains and trade rules. In: The shifting geography of global value chains: implications for developing countries and trade policy. s.l.: World Economic Forum, pp 16–17

[23]Available at: https://www.wto.org/english/news_e/spra_e/spra46_e.htm.

Baldwin R, Evenett S, Low P (2007) Beyond tariffs: multilaterising deeper RTA commitments. http://www.wto.org/english/tratop_e/region_e/con_sep07_e/baldwin_evenett_low_e.pdf. Accessed 11 Dec 2012

Borchert I, Gootiiz B, Mattoo A (2011) Services in Doha: what's on the table? In: Martin W, Mattoo A (eds) Unfinished Business? The WTO's Doha Agenda. CEPR and World Bank, London, pp 115–143

FAO (2011) The State of Food Insecurity in the World: how does international price volatility affect domestic economies and food security? Food and Agriculture Organization of The United Nations, Rome, 2011

Hermann M (2009) Food security and agricultural development in times of high commodity prices. Discussion paper no. No. 196, November 2009, UNCTAD, United Nations

Hoda A (2012) Global developments: a plurilateral services agreement? in trade policy and WTO Newsletter. http://www.icrier.org/pdf/april2012_wto_Newsletter.pdf

Hoda A, Gulati A (2007) WTO negotiations on agriculture and developing countries. Oxford University Press, New Delhi

Rodrik D (2008) Don't cry for Doha. http://rodrik.typepad.com/dani_rodriks_weblog/2008/07/dont-cry-for-doha.html. Accessed 10 Nov 2012

Is WTO Governed Trade Regime Sufficient for Export Growth?

Saikat Sinha Roy and Pradyut Kumar Pyne

Abstract The impact of WTO promulgated multilateral trade regime on world trade growth and welfare has remained an important issue since the formation of the institution in the mid-1990s. International trade economists have agreed to disagree on the subject. The finding of insignificant effect of WTO in promoting world trade, as empirically established by Rose (Am Econ Rev 94(1):98–114, 2004a), has been refuted by Subramanian and Wei (J Int Econ 71(1):151–175, 2007), Tomz et al. (Am Econ Rev 97(5):2005–2018, 2007), Helpman et al. (Q J Econ CXXIII(2), 2008), among others. The study investigates into whether WTO membership is sufficient for trade growth across countries. Using a gravity model framework and a dataset of 200 exporters and 234 importers, the study, controlling for trade facilitating infrastructure, time and country fixed effects along with other extended gravity variables, finds that WTO's trade regime, even though necessary, is not sufficient for growth in trade.

Keywords World trade organization · Multilateral trade · Export growth · Gravity model

JEL Classification F13 · F14 · F15

S. Sinha Roy (✉)
Department of Economics, Jadavpur University, Kolkata 700032, India
e-mail: saikat.sinharoy@gmail.com; ssroy@economics.jdvu.ac.in

P.K. Pyne
Department of Economic Affairs, Ministry of Finance, New Delhi 110001, India
e-mail: pradyut.pyne@gmail.com; pradyut.pyne@nic.in

© Springer India 2016
M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_9

1 Introduction

Market access and implementation of non-discriminatory trade policy measures using Most Favoured Nation (MFN) clause and national treatment are the key routes to promoting multilateral trade flows under World Trade Organization (WTO). While market access aims at removal and/or reduction in trade barriers, including tariff and non-tariff (NTB), a country can protect its sectors of interest through the provision of safeguards. Dispute settlement procedure of WTO monitors issues related to violation of agreement with regard to facilitating and maintaining multilateral relations. The formation of WTO as an institution during the Uruguay Round of multilateral trade negotiation, thus aims at ensuring free and fair trade multilaterally. The benefit from multilateral liberalization of trade in goods that was expected to accrue globally ranged between US$109 billion and US$510 billion by 2005.[1] The average annual growth of global exports of goods and services, based on World Economic Outlook (WEO) database, 6.1 % during 1985–1994 increased to 7.2 % during 1995–2007 from 6.1 % during 1985–1994. However, trade growth declined after the global economic crisis. This paper investigates into the impact of multilateral trade regime under WTO on trade growth across countries accounting for various other factors that determine trade.

The effect of WTO in multilateral trade promotion is one of the interesting issues considered by the international trade economists. The debate on the impact of WTO on world trade growth started with the findings of Rose (2004a) and the counter findings by Subramanian and Wei (2007). Rose (2004a) shows that GATT/WTO membership does not have any substantial effect on trade and only extension of Generalised Scheme of Preferences (GSP) from one country to another seems to have large positive impact. In another study, Rose (2004b) finds that members are often less open than non-members with regards to tariff and non-tariff barriers (NTBs) and concluded that there is no clear evidence that GATT/WTO accession or membership led to trade liberalization. Further, Rose (2004c) using pooled OLS regression argues that WTO/GATT membership increased trade volatility. In contrast, Subramanian and Wei (2007) provide evidence of positive but uneven impact of GATT/WTO on trade across the countries and sectors. The study argues that successive rounds of tariff liberalization under the aegis of GATT/WTO have served to increase industrial countries imports. Again, to explain the interesting finding of Rose (2004a), Tomz et al. (2007) argue that it is largely on account of mistaken country classification as WTO member and outsider. Considering non-member participation, the study finds that the GATT substantially increased trade of both formal member and non-member participants, relative to countries outside the agreement. In support of his earlier findings, Rose (2007) further shows that GATT had no substantial effect on aggregate trade and the same is true for both formal and informal membership. However, Helpman et al. (2008) argue that the commonly used gravity framework provides

[1]The estimate is based on the report prepared by GATT Secretariat on the results of the Uruguay Round of multilateral trade negotiations.

biased result on account of the non-consideration of zero bilateral trade pair in the dataset. It generalizes the Anderson and van Wincoop (2003) gravity framework in an international trade model by introducing firm heterogeneity, fixed trade cost and asymmetries in export flows in a bilateral pair, to control for potential biases. By using data on positive as well as zero trade flows between countries and using GATT/WTO membership variable as additional source of variation in export flows, the study shows that there is positive and very strong significant effect of GATT/WTO membership on bilateral trade flows, applying exporter and importer specific fixed effects. To reexamine the effect of GATT/WTO membership and overcome the risk of misspecification bias of parametric gravity methodology, Chang and Lee (2011) applied non parametric method and find large and significant trade promoting effect of GATT/WTO especially in case if both exporter and importer are member of WTO compared to the case when one country is WTO member. Eicher and Henn (2011) measure the WTO effect on trade flows by unifying the studies by Rose (2004a), Tomz et al. (2007), Subramanian and Wei (2007) by considering three types of omitted variable bias, namely, multilateral resistance, unobserved bilateral heterogeneity and individual PTA trade effects. The study finds that PTAs create trade, but unevenly across individual agreements. Countries with greater incentives to bargain for tariff reductions during WTO accession negotiations exhibit positive and significant WTO trade effects.

In the literature, the impact of GATT/WTO is also examined by decomposing trade into extensive and intensive margins.[2] Dutt et al. (2013) shows WTO membership has impacted on extensive margin of trade positively while it adversely affects intensive margin that in turn implies the significant role of WTO in reduction of fixed cost rather than variable cost of trade. The study suggests that WTO as an institution resolves uncertainties in the export market and promotes overall exports through diversification of export basket and market. On the whole, it is thus evident from the review of the existing literature that the impact of multilateral trade negotiation initiated by the formation of GATT/WTO on trade is far from resolved. The debate on the impact of WTO on trade centres around nuanced results largely on account of different methods of estimation adopted.

This paper investigates into the role of WTO and the impact of tariff and NTB liberalization on market access in merchandise trade using a larger time comparable dataset and accounting for the methodological biases, and thus adding to the existing literature. The trade cost factor that has not yet been taken account of in the literature is the role of trade facilitating infrastructure in determining trade flows. This is done while controlling for all other possible important determinants of multilateral trade in an extended gravity model framework. This paper helps to examine the extent of gains in terms of trade that have accrued to countries following WTO formation in 1995. However, the analysis that follows does not distinguish between developed, developing and least developed countries. In doing so, the following section

[2]For instance, Besedes and Prusa (2011) investigate into countries' export growth based on their performance at the extensive and intensive export margins.

provides a snapshot view on growth in world trade. Section 3 briefs on the gravity model and its extensions. In Sect. 4, methodological issues are delineated along with a brief illustration on data coverage and sources across countries. Estimation results are presented in Sect. 5. The paper concludes with summary of major findings highlighting the implications for policy in Sect. 6.

2 A Snapshot View on Growth in World Trade

To observe changes in world trade in goods and services, the period since 1980 has been taken into account to have a comparison between pre- and post-WTO periods. Growth of world trade in goods and services, based on IMF's World Economic Outlook (WEO) database, for the period since 1980 is presented in Fig. 1. It is evident that, with intermittent fluctuations, world trade growth averaged at 5.6 % during 1980–2010. After peaking in the first half of 2000s, both world trade and export growth declined since 2007. The sharpest fall in growth of global trade happened after 2009 following the financial crisis. Despite the downturn, average growth of global trade remained high during the post-WTO period (1995–2010). The average annual growth of trade in goods and services during pre (1980–1994) and post (1995–2010) WTO periods are 4.9 and 6.3 %, respectively, while the corresponding average annual growth of merchandise exports are 5.1 and 6.1 %, respectively. Along with observed higher average export growth in the post-WTO period, there has been lowering of fluctuations in growth. This is evident in Fig. 2, the coefficient of variation of export growth across countries is found to have

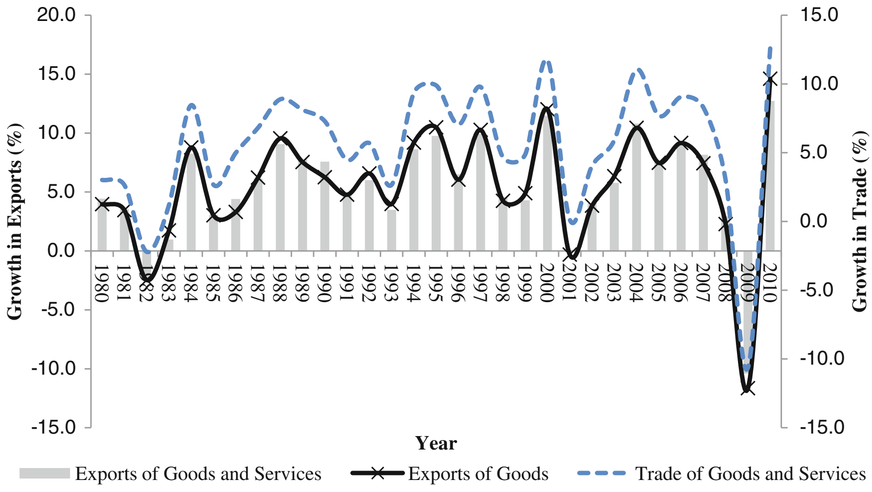


Fig. 1 Growth in world trade in goods and services

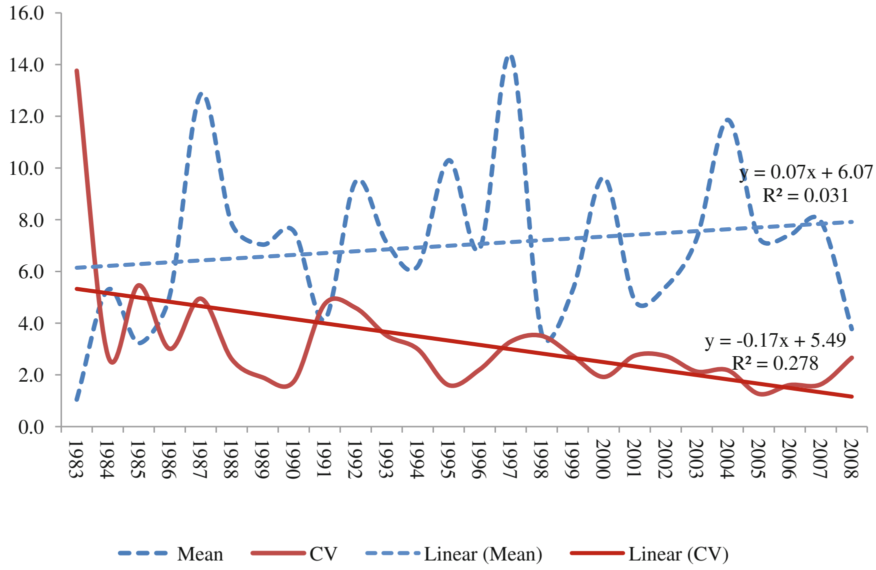


Fig. 2 Coefficient of variation of export growth across countries

declined over time. Having observed average global trade growth and its variations, it is important to understand the pattern of export growth across countries.

2.1 Cross-Country Export Growth

Variations in growth of exports across countries are analyzed using multiple Box-and-Whiskers plot of average growth rates of exports across countries over six subperiods during the 30-year period since 1981. Observations with growth above 40 % (abnormal outlier) are not considered as it is rare. A marginal improvement in mean growth rate is found during 1996–2000 and 2001–05 compared to the earlier subperiods (see Fig. 3). The marginal decline in mean growth during 2006–10 might be on account of the global financial crisis since 2007–08. It is evident from the plot in Fig. 3 that the heights of the boxes during the post-WTO period (1996–2000, 2001–2005, and 2006–10) are comparatively shorter than that for pre-WTO period (1981–85, 1986–90, and 1991–95), which points to the decline in the variation of export growth rates across countries. In addition, the number of outliers is lesser during the post-WTO period. This is indicative of mobility of countries and plausible convergence of export growth across countries during the post-WTO period. The following subsection provides a brief on mobility of countries in terms of export growth rates in the post-WTO period.

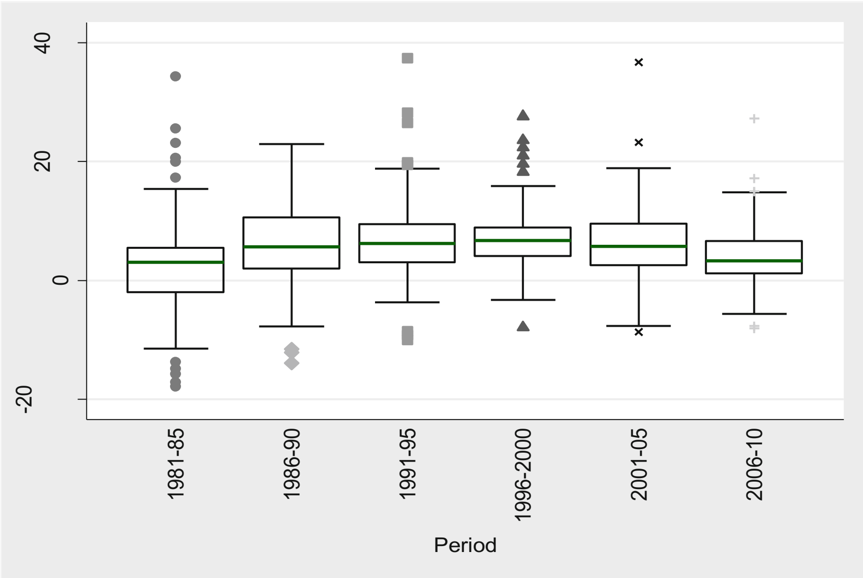


Fig. 3 Distribution of export growth across countries. *Source* WEO database, IMF

2.1.1 Mobility Analysis

To study the extent of movement with respect to export growth across countries, an analysis has been carried out using mean and variance of annual average export growth rates. To justify the widely held assumption of the increase in the opportunities of moving up in terms of export growth after the formation of WTO, both member and non-member countries of WTO are considered in the analysis. There are differences in the number of countries in each period with an increase in the coverage in the dataset than in the earlier years.

On the basis of the IMF's WEO data, Table 1 shows mobility of countries based on their own export growth rate for the period before and after the formation of WTO. It has been noticed that prior to the formation of WTO around 90 % of the countries registered export growth within the (Mean ± 1 s.d.) band of growth. Even though the proportion of countries increased to around 93 % just after the formation of WTO, relatively larger number of countries, those which belong outside the (Mean ± 1 s.d.) range, are maintaining their position in their respective regions (Table 1). This is contrary to the perception that lesser percentage of countries inside the (Mean ± 1 s.d.) range are maintaining their earlier period's position during the post-WTO period. It can therefore be argued that even though there is a decline in the variations of export growth rates across countries, WTO formation has not ensured better mobility of the countries to achieve the mean

Table 1 Movement of the countries over the periods

		Above the (mean + 1 s.d.) band	Inside the (mean ± 1 s.d.) range	Below the (mean – 1 s.d.) band
I: Movement of countries (%) from period A to period B		*Period B (1991–95)*		
Period A (1986–90)	Above the (mean + 1 s.d.) band	20.0	80.0	0.0
	Inside the (mean ± 1 s.d.) range	6.6	90.9	2.5
	Below the (mean – 1 s.d.) band	0.0	100.0	0.0
II: Movement of countries (%) from period B to period C		*Period C (1996–2000)*		
Period B (1991–95)	Above the (mean + 1 s.d.) band	27.3	63.6	9.1
	Inside the (mean ± 1 s.d.) range	5.1	93.5	1.4
	Below the (mean – 1 s.d.) band	12.5	75.0	12.5
III: Movement of countries (%) from period D to period E		*Period E (2001–05)*		
Period D (1996–00)	Above the (mean + 1 s.d.) band	8.3	91.7	0.0
	Inside the (mean ± 1 s.d.) range	6.6	86.2	7.2
	Below the (mean – 1 s.d.) band	33.3	50.0	16.7

Source Own calculation on the basis of data from WEO, IMF

growth rate in each subperiod. On the whole, even if average growth of exports improved after the formation of WTO in 1995, it is unlikely that all countries have performed equally well. At this juncture it is important to understand whether such improvements in trade flows are consequent upon the formation of WTO, with a multilateral trade order in place and the likely improvements in market access. In what follows is an understanding, using an econometric exercise, of the role of WTO in the determination of cross-country trade flows.

3 Gravity Model in Trade Flow Estimation and Its Extensions

In applied international trade modelling, the use of gravity model is widespread. Important studies, including Linneman (1966), Bergstrand (1985, 1989), etc., use the gravity model essentially deriving the gravity equation from a general equilibrium framework. In this study the gravity model is used to explain the factors underlying bilateral trade flows in a multilateral framework.

The gravity model, in its basic form assumes that economic mass and geographical distance among trading nations are key to the bilateral trade flows (gravity model assume perfect international product substitutability). Following Newton's law of gravitation, the basic form of gravity model is as

$$X_{ij} = f\left(Y_i, Y_j, \tau_{ij}\right), \text{with } \delta X/\delta Y > 0 \text{ and } \delta X/\delta \tau < 0 \tag{1}$$

where, X_{ij} represents volume of trade flow (export or import or total trade volume) from country i to j, Y is the gross domestic product and τ represents trade cost between the countries.

Geographical distance between the countries can be an observable proxy for trade cost. The gravity framework, thus, shows that larger country pairs are expected to trade more, while more distant country pairs trade less, perhaps on account of higher transport cost. Some studies (Brun et al. 2005; Coe et al. 2007) have established that distance as a factor determining bilateral trade has lost its importance as a proxy for trade cost following globalization, but Carrere and Schiff (2005) show increasing importance of distance in determining trade flows. Further, the basic gravity model does not consider the impact of reduction of trade costs following trade agreements that a country engages into. This is on account of the fact that such agreements are found to significantly lower trade costs between two partner countries while having no impact between non-member trade partners, i.e.

$$\delta X_{ij}/\delta \tau_{ik} = 0 \tag{2}$$

where i and k are members of a trade agreement but j is not. This framework thus ignores the effects of trade creation and trade diversion.

Anderson and van Wincoop (2003) augment the basic gravity model by incorporating multilateral resistance parameters (both external and internal). These resistance factors actually capture the dependence of trade flows between two countries on trade costs in all possible ways. Common factors influencing trade costs in the literature are geographical distances between the trade partners, sharing of common land border, having common official language, past colonial relationship between the trade partners, and the case of colonization by the same power, etc. The model considers the effect of changes in trade costs among a pair of bilateral trade partner on trade flows of all other partners because of the change in relative price. Hummels (2001) measures trade costs that affects trade flows in terms of explicitly measured cost (tariff and freight), costs associated with common proxy variables (distance, adjacency, linguistic similarities), and an unmeasured part.

In the empirical international trade literature, the gravity model has been used in the context of trade flows in regional trade agreements (Frankel 1997; De 2004; Baier and Bergstrand 2007; Jayasinghe and Sarkar 2008; Kepaptsoglou et al. 2010; Egger et al. 2011; Eicher and Henn 2011; Martin et al. 2012 etc.); currency unions (Rose 2000, 2004a, c; Glick and Rose 2002); multilateral trade (Subramanian and Wei 2007; Tomz et al. 2007; Dutt et al. 2013), among others. From a review of these studies it is evident that the gravity model, when suitably adjusted, is appropriate in estimating bilateral trade flows in a multilateral framework.

4 Model and Econometric Methodology

4.1 Possible Determinants of Export Flows

The literature on determinants of multilateral trade flows, accounting for the impact of formation of WTO use gravity model. Apart from extended gravity variables like distance and per capita GDP of trade partners, colonial history, linguistic similarities, contiguity, etc., country specific fixed effects are accounted for following the suggestions of Anderson and van Wincoop (2003). Following Helpman et al. (2008), the study will consider all possible bilateral pairs in the econometric exercise. Further, to examine the effects of WTO and other policy variables, some mutually exclusive and exhaustive dummies are taken into account to separate out the WTO membership effect. This follows Dutt et al. (2013). This study adds to the literature by taking into account variables measuring trade costs such as infrastructure, tariff and non-tariff restrictions along with the variables generally considered in the extended gravity model framework.

It is established in the existing literature that infrastructure is critical for economic growth and open economies can provide better infrastructure than its closed counter parts (WTO 2004; De 2004; Chakravorty and Mazumdar 2008). As an important determinant of transport cost, infrastructure is considered to be directly related to firm productivity, diversification in production and expansion in trade, increase in output and economic growth (Munnell 1992; Easterly and Rebelo 1993; Limao and Venables 2001; De 2004, 2006). However, the effective rate of protection in international trade flow provided by transport cost, in many of the countries, is considerably higher than that by tariff.[3] Further, the lack of quality infrastructure creates obstacles to the supply chain of production and trading of final goods as well.[4] For Asian countries, De (2004) finds the significant role of transaction cost and infrastructure in explaining variation in trade flows in a gravity

[3]WTO (2004).

[4]Poor infrastructure has been identified as one of the responsible factor that restricts participation of the developing countries in the global value chain (WTO 2014).

model framework. Chakravorty and Mazumdar (2008) develop a model of trade in which provision of infrastructure makes the domestic firm more productive and enables to capture market share from a foreign firm in both the domestic and export markets. Infrastructure is thus treated as a complementary factor to the trade policies for multilateral trade expansion and, hence, is included in the model.

4.1.1 Measure of Infrastructure

For the measurement of trade facilitating infrastructure (TFI) variable, principal component analysis (PCA) is used in this study. The variables used for measuring TFI across countries during 1995–2010 include volume of air freight measured in metric tonnes times kilometre travelled, flow of port containers from land to sea transport modes and vice versa in a standard size container (20-foot equivalent units), road density measured by the ratio of the length of country's total road network to the country's land area, volume of goods transported by railways measured in metric tonnes times kilometre travelled, internet users per 100 people, and telephone lines per 100 people. Data on these variables are collected from World Bank's World Development Indicator (WDI) database at the country level. For the sake of brevity, the weighting diagram of the variables included in the infrastructure index is not provided herewith.

4.1.2 The Econometric Model

In order to analyze the impact of formulation of WTO on export flows, following Rose (2004a), Subramanian and Wei (2007), Tomz et al. (2007), Dutt et al. (2013), De (2004) among others, the following gravity equation is used

$$X_{ij} = \beta_0 \cdot D_{ij}^{\beta_1} \cdot Y_i^{\beta_2} \cdot Y_j^{\beta_3} \cdot \lambda_{ij}^{\beta_4} \cdot \gamma_{ij}^{\beta_5} \cdot \mathrm{TFI}^{\beta_6} \tag{3}$$

where i and j are trade partner countries. The gravity variables are defined as follows:

X_{ij} is the flow of export from ith country to jth country;
D_{ij} is the distance between source and destination country;
Y_i is the per capita GDP of the source country;
Y_j is the per capita GDP of the destination country;
λ_{ij} represents the set of time invariant trade influencing geographical, cultural, and historical factors like linguistic similarities, colonial relationship, sharing common land border, etc., among trading countries. Linguistic similarity (*Common language*) is a binary variable which is unity if both the trading partners have common official language. Again colony (*Colonial link*) is a binary variable which is unity if i ever colonized j or vice versa, and contiguity (*Border*) is a binary variable which is unity if i and j share a land border;

γ_{it} represents the set of country specific issues like WTO membership, membership of a Regional Trade Agreement (RTA), beneficiary of a Preferential Trade Arrangements (PTA) including offering of GSP, etc., and

TFI is the measure of trade facilitating infrastructure variable for the trading nations

In log-linear form, the Eq. (3) can be expressed as

$$\ln X_{ijt} = \beta_0 + \sum_i E_i + \sum_j M_j + \sum_t T + \beta_1 \ln D_{ij} + \beta_2 \ln Y_{it} + \beta_3 \ln Y_{jt} + \beta_4 \lambda_{ij} + \beta_5 \gamma_{ijt} + \beta_6 \mathrm{TFI}_t + \mu \tag{4}$$

Here t represents year, E_i, M_j and T_t represent the country and year specific dummies to have country a year specific fixed effects, and μ is the residual disturbance term assumed to be well behaved. β_i's are the respective coefficients of the variables used in the regression.

In order to examine the effect of tariff and non-tariff barriers on multilateral trade flows, a measure of tariff and non-tariff barriers is considered for the estimation in a stepwise manner. First, to analyze the sole effect of these trade restricting barriers a set of estimations is carried out by dropping the WTO membership and related dummy variables from the Eq. (4). In that case, Eq. (4) can be represented as

$$\ln X_{ijt} = \beta_0 + \sum_i E_i + \sum_j M_j + \sum_t T + \beta_1 \ln D_{ij} + \beta_2 \ln Y_{it} + \beta_3 \ln Y_{jt} + \beta_4 \lambda_{ij} + \beta_5 \delta_{ijt} + \beta_6 \mathrm{TFI}_t + \mu \tag{5}$$

where δ_{ij} represent time variant trade restricting policy instruments like tariff and non-tariff barriers. The second step involves the estimation of the combined effect of tariff and non-tariff barriers along with WTO membership and related policy variables.

Separating Out the Effect of WTO

Following Subramanian and Wei (2007), Eicher and Henn (2011) and the decomposition of trade preferences as carried out by Dutt et al. (2013), the paper also accounts for WTO membership by the trading nations, RTA and PTA using a set of mutually exclusive and exhaustive dummies to estimate the pure WTO effect. The dummies are as follows

1. **Dummy for pure WTO effect (D1)**: The dummy will take the value 1, if both source and destination countries are members of WTO but they do not belong to a RTA and the destination does not offer any preferential arrangement to the exporter, and zero otherwise.

2. **Dummy for pure RTA effect (D2)**: The dummy will take the value 1, if both the trading partners are members of a common RTA, but at least one of them is not member of WTO and the destination does not offer any preferential arrangement to the exporter, and zero otherwise.
3. **Dummy for pure PTA effect (D3)**: The dummy will take the value 1, if the importer extends preferential arrangements to the exporter but at least one of them is not member of WTO and they do not belong to any RTA, and zero otherwise.
4. **Dummy for effect without PTA (D4)**: The dummy will take the value 1, if both source and destination countries are members of WTO and the trading partners are members of a common RTA, but the destination does not offer any preferential arrangement to the exporter, and zero otherwise.
5. **Dummy for effect without RTA (D5)**: The dummy will take the value 1, if both origin and source countries are member of WTO and importer extends preferential arrangement to the exporter but they do not belong to any RTA, and zero otherwise.
6. **Dummy for effect without WTO (D6)**: The dummy will take the value 1, if importer extends preferential arrangement to the exporter and both the countries belong to same RTA but at least one of them is not member of WTO, and zero otherwise.
7. **Dummy for having all three effects**: The dummy will take the value 1, if both origin and source countries are members of WTO, the trading partners are member of common RTA and importer extends preferential arrangements to the exporter, and zero otherwise.

The estimated coefficient of D1 will thus give the pure effect of WTO on export flows. The asymmetric impact of WTO, if any, can also be examined by separate set of estimations for developed, developing and least developed economies, which is however outside the scope of this paper.

4.2 Data Coverage and Sources

The econometric estimation considers a large dataset of about 5.2 lakh data points with 200 source countries and 234 partner countries for the period 1995–2010. The study includes all countries, developed, developing and least developed. For estimation of the model, data on merchandise export value (in US dollar term) are collected for all countries from *World Integrated Trade System (WITS)* database published by the World Bank. The export data are based on HS (1989–92) classification. Data on per capita gross domestic product (PCGDP) at constant 2000 US $ prices are collected form *World Development Indicator (WDI)* database. Data on distance between the countries (in kilometre) are availed from the *CEPII* database.[5] Other gravity variables including linguistic similarities, colony, contiguity, etc., are also taken from the *CEPII* database. WTO membership is used as determining factor for export flows. Information on WTO membership, the Preferential Trading Arrangements (PTA) along with Generalised Scheme of Preferences (GSP), and Regional Trade Agreement (RTA) information are collected from WTO database (www.wto.org). Data on ad-valorem tariff and non-tariff costs of bilateral trading pairs are collected from ESCAP database on comprehensive trade cost.[6]

5 The Estimation Results

The estimation has been done considering all possible trade pairs in two sets of model specification: (i) without tariff and non-tariff measures, and (ii) with tariff and non-tariff measures. As mentioned earlier, the estimations are controlled for either time fixed effects, exporter–importer fixed effects or both. The baseline estimation is done without any control. Estimation results of models M(1) to M(8) are presented in Table 2. Results without any fixed effects are reported in model 1 and 2. Year specific and exporter–importer specific dummies are used stepwise and presented in models 3 to 8. The F-statistics of all the regression equations, as reported in Table 2, show that the estimations are significant at 1 % level. Further,

[5]As per the CEPII database, distance between two countries is calculated using the great circle formula, which uses latitudes and longitudes of the most important cities/agglomerations (in terms of population). In most cases the main city is the capital of the country, but for very few countries the capital is not populated enough to represent the economic center of the country.

[6]ESCAP Comprehensive Trade Cost (CTC) database measure the bilateral CTC between country *i* and *j*, following Chen and Novy (2009) as

$$\mathrm{ctc}_{ij} = \sqrt{\left(t_{ij}.t_{ji}/t_{ii}.t_{jj}\right)} = \left(x_{ii}.x_{jj}/x_{ij}.x_{ji}\right)^{\frac{1}{2(\sigma-1)}}$$

where, t_{ij} (t_{ji}) represents international trade costs from country $i(j)$ to country $j(i)$. t_{ii} (t_{jj}) denotes international trade costs of country $i(j)$. x_{ij} (x_{ji}) denotes international trade flow from country i (j) to country $j(i)$. σ denotes elasticity of substitution between goods.

Bilateral tariff cost measure in this database is defined as

$$T_{ij} = \sqrt{\left(1+\mathrm{tariff}_{ij}\right)\left(1+\mathrm{tariff}_{ji}\right)}$$

where, tariff_{ij} (tariff_{ji}) is the simple average effective import tariff imposed by country $i(j)$ on country $j(i)$.

All additional trade cost other than tariff costs involved in trading goods bilaterally rather than domestically are calculated as non-tariff trade cost as

$$\mathrm{ntctc}_{ij} = \frac{\mathrm{ctc}_{ij}}{T_{ij}}$$

Table 2 Gravity model estimation: exports from all countries

Variables	Without fixed effect		With fixed effect					
			Without tariff and non-tariff			With tariff and non-tariff		
	M(1)	M(2)	M(3)	M(4)	M(5)	M(6)	M(7)	M(8)
LN (GDP_Exporter)	0.47***	0.07***	0.38***	1.65***	1.16***	0.04***	1.17***	0.89***
	(0.01)	(0.01)	(0.01)	(0.05)	(0.06)	(0.01)	(0.06)	(0.06)
LN (GDP_Importer)	0.20***	−0.02**	0.11***	1.07***	0.77***	−0.06***	1.29***	1.01***
	(0.01)	(0.01)	(0.01)	(0.04)	(0.05)	(0.01)	(0.05)	(0.06)
LN (Distance)	−0.96***	0.51***	−1.06***	−1.72***	−1.72***	0.39***	−0.16***	−0.16***
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
LN (TFI_Exporter)	0.27***	0.16***	0.39***	0.01***	0.00*	0.20***	0.00	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LN (TFI_Importer)	0.18***	0.03***	0.27***	0.01***	0.00*	0.06***	0.00	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Common language	0.12**	−0.46***	0.37***	0.77***	0.78***	−0.36***	0.08***	0.08***
	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.02)	(0.02)
Border	2.05***	0.07	1.88***	0.91***	0.92***	0.08	−0.23***	−0.23***
	(0.12)	(0.10)	(0.12)	(0.12)	(0.12)	(0.10)	(0.05)	(0.05)
Colonial link	2.32***	0.36***	2.02***	1.04***	1.04***	0.34***	−0.02	−0.02
	(0.13)	(0.10)	(0.12)	(0.11)	(0.11)	(0.10)	(0.04)	(0.04)
D1: pure WTO effect	1.97*	0.49	1.86	0.91	0.95	0.45	−0.32	−0.31
	(1.01)	(0.45)	(1.16)	(0.95)	(0.92)	(0.58)	(0.33)	(0.31)
D2: pure RTA effect	0.73***	0.25***	0.42***	0.52***	0.50***	0.12*	0.15***	0.12***
	(0.09)	(0.07)	(0.08)	(0.07)	(0.07)	(0.07)	(0.04)	(0.04)
D3: pure PTA effect	0.72***	0.20**	0.58***	0.06	0.05	0.18**	0.16***	0.16***
	(0.11)	(0.08)	(0.10)	(0.08)	(0.08)	(0.08)	(0.06)	(0.06)

(continued)

Table 2 (continued)

Variables	Without fixed effect		With fixed effect					
			Without tariff and non-tariff			With tariff and non-tariff		
	M(1)	M(2)	M(3)	M(4)	M(5)	M(6)	M(7)	M(8)
D4: effect without PTA	1.36***	0.32***	1.27***	0.26***	0.25***	0.30***	−0.05***	−0.05**
	(0.06)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.02)	(0.02)
D5: effect without RTA	0.80***	−0.15***	0.97***	0.31***	0.29***	−0.09*	0.08***	0.08***
	(0.07)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.03)	(0.03)
D6: effect without WTO	−0.34	−0.15	−0.05	−0.06	−0.10	−0.10	0.17	0.14
	(0.29)	(0.18)	(0.26)	(0.19)	(0.19)	(0.18)	(0.12)	(0.12)
D7: effect of all	0.50***	−0.07	0.73***	0.41***	0.37***	−0.02	0.04	0.04
	(0.14)	(0.07)	(0.13)	(0.09)	(0.09)	(0.07)	(0.05)	(0.05)
Tariff		−6.98***				−7.53***	−6.48***	−6.36***
		(0.25)				(0.24)	(0.13)	(0.13)
NTBs		−4.89***				−4.66***	−3.97***	−3.98***
		(0.04)				(0.04)	(0.02)	(0.02)
Constant	6.32***	28.67***	8.79***	−8.74***	−3.40***	29.46***	11.97	15.85
	(0.28)	(0.30)	(0.26)	(1.12)	(1.19)	(0.29)	.	.
Time FE	No	No	Yes	No	Yes	Yes	No	Yes
Exporter FE	No	No	No	Yes	Yes	No	Yes	Yes
Importer FE	No	No	No	Yes	Yes	No	Yes	Yes
Observations	249793	147132	249793	249793	249793	147132	147132	147132
Adjusted R^2	0.386	0.703	0.452	0.716	0.717	0.712	0.857	0.857

Standard errors in parentheses, $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

it is important to note that the adjusted R^2 improves (from 70 to 85 %) when tariff and non-tariff measures are taken into account along with time and country specific fixed effects. All these indicate that the estimation model, the extended gravity model, fits the data well.

Estimation results show that the basic gravity variables are significant at 1 % level with expected signs in most cases. The economic size of trading countries, measured by the per capita GDP, has positive impact on exports in almost all cases with larger impact of exporting country's per capita GDP than that of the destination country. As expected, distance as an indicator of trade cost impact export flows inversely. This result on trade cost, measured in terms of distance, shows countries which are geographically close have lower trade costs leading to higher trade flows between them.

The extended gravity variables, for instance, common land border, colonial history and common language (in most cases) have positive and significant impact on export flows. The exception to this result is found: common land border in the models with tariff and non-tariff barriers and in the presence of country specific fixed effects has negative and significant impact on export flows. Trade facilitating infrastructure of both trading partners is found to be a positive and significant determinant of export flows. Infrastructure of both trading countries is found to be an important determinant of export flows because it ensures reduction in transactions costs for exporters. The impact of trade facilitating infrastructure, as expected, is stronger for the exporting country than for partner countries. On the other hand, lack of trade facilitating infrastructure increases transactions costs to trade thereby reducing trade potential.

However, distance between two trade partners and their respective infrastructure do not take into account the entire range of trade costs. The literature shows that tariff as well as the entire gamut of non-tariff barriers to trade form an increasing proportion of trade costs. In the regression estimation, thus trade costs arising out tariff and non-tariff barriers are taken into account. The results presented in Table 2 (models 6–8) show that tariff and non-tariff barriers have a significant inverse relationship with bilateral export flows. The elasticities are found to high. Lowering of these trade barriers improve trade flows in a significant way. However, the presence of these two variables in the regressions reduces the explanatory power of the infrastructure and relegates the infrastructure variable to insignificance in export determination.

Apart from basic and augmented gravity variables, WTO membership of both exporter and importer countries is found have significant impact on export flows between a pair of countries. Such WTO membership does not necessarily separate out the trading countries membership in an RTA or a PTA. The pure policy effects like PTA and RTA are found to be positive and significant. The effect of WTO without either RTA or PTA is also positive and significant in all the cases. However, the effect of PTA and RTA without WTO is insignificant. After netting out membership in a RTA or a PTA, pure WTO membership is taken as an explanatory variable. It is found that pure WTO membership has insignificant impact on export flows except in model 1. This result confirms the findings of Rose

(2004a). Even though the impact of WTO membership along with that of RTA and PTA is positive significant, pure WTO effect has insignificant impact on bilateral trade flows in a multilateral framework.

On the whole, the basic and extended gravity variables remain important as determinants of trade flows in a trade regime prevalent after WTO formation. Trade facilitating infrastructure, tariff and non-tariff barriers by way of reducing transactions costs, are found to be significant determinants of trade flows. The effect of WTO membership is important in determining trade flows, but pure WTO membership is not found to significantly impact on export flows between countries. Here it will be interesting to understand the impact of these measures across different country groups, which is outside the scope of this study.

6 Conclusions

The issue of the role of WTO formation in export promotion has been debated extensively by Rose (2004a, b, c, 2007), Subramanian and Wei (2007). Many studies delved into this debate, raised many methodological and other issues, but the debate remained inconclusive. This study adds to the debate in the following ways: (i) increasing the coverage of trade pairs (source and destination countries), (ii) separating out the effect of WTO, PTA and RTA following the decomposition done by Dutt et al. (2013), (iii) including trade facilitating infrastructure as an explanatory variable, and (iv) using tariff and non-tariff measures in the gravity model. In the econometric estimation, the study controls for time and country specific fixed effects.

The estimation results show that basic as well as extended gravity variables are significant in determining trade flows between pairs of countries. With respect to per capita GDP, it is found that the export is more driven by the supply side. Distance is significantly explaining exports, which is contrary to the perception of significant reduction in distance between countries during globalization. Other extended gravity variables such as linguistic similarities and colonial history impact on exports significantly while sharing common land border does not induce increase in exports between country pairs in a significant way. Infrastructure in facilitating exports in both source and destination country is playing an important role. Specifically, for developing countries the role of trade facilitating infrastructure is expected to be crucial in export promotion given the infrastructure deficiency in these countries. Tariff and non-tariff barriers also strongly determine bilateral exports.

Form the estimation results it is found that WTO membership along with membership in a PTA or/and RTA is found to be significant, while a pure WTO membership or a pure PTA/RTA membership comes out to be insignificant. Even though WTO membership is important, a multilateral trade regime governed by WTO is not sufficient for trade promotion. It is instructive to note that regional and/or preferential membership are important in trade promotion along with other complementary factors including lower tariff and non-tariff barriers and trade facilitating infrastructure. The effect of WTO membership and other factors can however vary with the level of development of the countries engaged in international trade.

References

Anderson K, van Wincoop (2003) Gravity with gravitas: a solution to the border puzzle. Am Econ Rev 93(1):170–192

Baire SL, Bergstrand JH (2007) Do free trade agreements actually increase members international trade? J Int Econ 71(1):72–95

Bergstrand JH (1985) The gravity equation in international trade: some microeconomic foundation and empirical evidence. Rev Econ Stat 67(3):474–481

Bergstrand JH (1989) The generalized gravity equation, monopolistic competition, and the factor proportions theory in international trade. Rev Econ Stat 71(1):143–153

Besedes T, Prusa TJ (2011) The role of extensive and intensive margins and export growth. J Dev Econ 96(2):371–379

Brun JF, Carrere C, Guillaumont P, Melo JD (2005) Has distance died? evidence from a panel gravity model. World Bank Econ Rev 19(1):99–120

Carrere C, Schiff M (2005) On the geography of trade: distance is alive and well. Revue économique 56(6):1249–1274

Chakravorty U, Mazumdar J (2008) Openness, lobbying, and provision of infrastructure. South Econ J 74(4):1149–1166

Chang P, Lee M (2011) The WTO trade effect. J Int Econ 85(1):53–71

Chen N, Novy D (2009) International trade integration: a disaggregated approach. CEP Discussion Papers dp0908, Centre for Economic Performance, LSE

Coe DT, Subramanian A, Tamirisa NT (2007) The missing globalization puzzle: evidence of the declining importance of distance. IMF Staff Papers 54(1):34–58

De P (2004) Transaction costs as barriers to economic integration in Asia: an empirical exploration. RIS Discussion Paper no. 77. Research and Information System for Developing Countries, New Delhi

De P (2006) Why trade cost matter? ARTNeT working paper series no. 7. UNESCAP-Asia-Pacific Research and Training Network on Trade, Bangkok

Dutt P, Mihov I, Zandt TV (2013) The effect of WTO on the extensive and the intensive margin of trade. J Int Econ 91(2):204–219

Easterly W, Rebelo S (1993) Fiscal policy and economic growth: an empirical investigation. J Monetary Econ 32(3):417–458

Egger P, Larch M, Staub KE, Winkelmann R (2011) The trade effects of endogenous preferential trade agreements. American Econ J: Econ Policy 3(3):113–143

Eicher TS, Henn C (2011) In search of WTO trade effects: preferential trade agreements promote trade strongly', but unevenly. J Int Econ 83(2):137–153

Frankel J (1997) Regional trading blocs in the world economic system. Institute for International Economics, Washington D.C

Glick R, Rose AK (2002) Does a currency union affect trade? the time series evidence. Eur Econ Rev 46(6):1125–1151

Helpman E, Melitz M, Rubinstein Y (2008) Estimating trade flows: trading partners and trading volumes. Q J Econ CXXIII(2):441–487

Hummels D (2001) Towards a geography of trade costs. Working paper, Purdue University, West Lafayette

Jayasinghe S, Sarkar R (2008) Effects of regional trade agreements on trade in agrifood products: evidence from gravity modelling using disaggregated data. Rev Agric Econ 30(1):61–81

Kepaptsoglou K, Karlaftis MG, Tsamboulas D (2010) The gravity model specification for modelling international trade flows and free trade agreement effects: a 10-year review of empirical studies. Open Econ J 3:1–13. http://benthamopen.com/contents/pdf/TOECONSJ/TOECONSJ-3-1.pdf

Limao N, Venables AJ (2001) Infrastructure, geographical disadvantage, transport cost and trade. World Bank Econ Rev 15(3):451–479

Linneman H (1966) An econometric study of international trade flows. North Holland Publishing, Amsterdam

Martin P, Mayer T, Thoenig M (2012) The geography of conflicts and regional trade agreements. Am Econ J: Macroecon 4(4):1–35

Munnell AH (1992) Infrastructure investment and economic growth. J Econ Perspect 6(4): 189–198

Rose AK (2000) One money one market: the effect of common currencies on trade. Econ Policy 15(30):7–45

Rose AK (2004a) Do we really know that the WTO increases trade? Am Econ Rev 94(1):98–114

Rose AK (2004b) Do WTO members have a more liberal trade policy? J Int Econ 63(2):209–235

Rose AK (2004c) Does the WTO make trade more stable? NBER working paper series no. 10207. National Bureau of Economic Research, Cambridge MA

Rose AK (2007) Do we really know that the WTO increases trade? reply. Am Econ Rev 97(5):2019–2025

Subramanian A, Wei SJ (2007) The WTO promotes trade, strongly but unevenly. J Int Econ 71(1):151–175

Tomz M, Goldstein JL, Rivers D (2007) Do we really know that the WTO increases trade? comment. Am Econ Rev 97(5):2005–2018

WTO (Various Years) World trade report, various years. World Trade Organisation, Geneva

Part III
Issues in Trade, Trade Policy and Development

Export Performance in Textile and Garments with China as a Competitor: An Analysis of India's Situation from the Perspective of Structure-Conduct-Performance Paradigm

Sarmila Banerjee, Sudeshna Chattopadhyay and Kausik Lahiri

Abstract Since the beginning of this millennium both India and China are recognized all over the world as two emerging economies. Both of them are members of the WTO and considered as two major competing players for textile and garments (T&G) products. Over the last 18 years the share of India's T&G export in the world market has gone up from 2.87 to 4.14 % and during the same period the increase experienced by China is from 12.97 to 39.08 %. In general China reveals stronger comparative advantage compared to India in almost all component groups of T&G and the situation has not altered much after the abolition of MFA in 2005. It has been shown in the analysis that in the international platform there is no definite evidence of China receiving any favorable treatment in terms of market access to explain this performance differential. So, the root cause has been searched in the domestic policies that shape the structure of the industry and influence its performance through creating incentives for different types of conducts. While India is specializing in the relatively high-value products in her niche markets, China is entering in large scale in all varieties of T&G products in both traditional as well as newly explored markets. It is seen that while Chinese T&G sector is expanding by capitalizing the economies of large-scale production, the market served by India relies mostly on its scope economy.

S. Banerjee (✉)
University of Calcutta, Kolkata, India
e-mail: sbeco@caluniv.ac.in

S. Chattopadhyay
Barasat Government College, Kolkata, India
e-mail: sudeshna.chattopadhyay@gmail.com

K. Lahiri
Surendranath College, Kolkata, India
e-mail: alklahiri@yahoo.co.in

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_10

Keywords Post-MFA · RCA · CMS · Finger-Kreinin similarity index · Scale economy · Scope economy · Performance in value chain

JEL Classifications F14 · L25 · L67

1 Context

This paper will examine why India and China, two emerging large economies, are performing so differently in the World market for Textile and Garments (T&G) over the post-WTO, post-MFA period, when the forces of competition are supposedly ensuring *level playing field* for all the major players. Between 1995 and 2012, the share of India in the total world export of T&G has gone up from 2.87 to 4.46 % (less than twofold) and over the same period the increase in China's share was from 12.97 to 39.09 % (more than threefold). The T&G industry is a very important export sector in both these countries and up to 2004 both faced quota restrictions in the markets of developed countries due to the Multi Fiber Agreement (MFA). The MFA was initially designed to protect the industry in the developed countries from cheap imports from the developing countries and allowed application of selective quota restrictions whenever these inflows threatened to cause serious damage to the industry of the importing country. However, under the aegis of the World Trade Organization (WTO) all the member countries are sharing more or less similar regulatory support from this supranational institution in carrying out international trade and in settling disputes arising there from. The MFA was thus, a major departure from the basic GATT rules and particularly from the principle of nondiscrimination. On 1 January 1995, the MFA was replaced by the Agreement on Textiles and Clothing (ATC) as an offshoot of the initiatives taken by the WTO, which set out a transitional process for the ultimate removal of these quotas.

The ATC aimed to liberalizes trade in textile and garments and set out a four stage quota liberalization schedule. In each phase a specified portion of the quota based on importing country's 1990 import volume was removed. The first phase of quota removal (between 1995 and the end of 1997) aimed at elimination of quantitative restrictions from a minimum of 16 % of import, followed by batches of a minimum of 17, 18 %, and finally 49 % (the remaining categories). The ATC was thus a time limiting agreement and ceased to exist from 1 January 2005. The products that were phased out in each round had to include goods from all major textile and clothing segments that is, yarn, fabrics, made-up textile products, and clothing. Since, the importing countries had the freedom to choose the products that they could liberalize in the first three phases, they chose those products for which quotas were least binding for the exporting countries. Thus, the process of implementation of the phasing out of the system of the MFA quotas was slow with most of the integration taking place in the final phase. Since, India was a member of the WTO since its inception in January 1995, she was eligible for quota reductions from the very beginning of the first phase. China on the other hand being outside the WTO was not eligible for the first two phases of quota removal. China became eligible for quota reductions only after it joined the WTO in December 2001. The Agreement whereby China earned full membership status of the WTO allowed other member countries to take special safeguards[1] against China until the end of 2008. After the expiry of the MFA quotas in 2005 these safeguards against China were used by the European Union (EU) till the end of 2007 and by the United States of America (USA) till the end of 2008, to protect their domestic interest against excessive Chinese invasions. Compared to India–China not only faced more stringent quota restrictions in the international market at any given point in time, China was subjected to a much extended quota regime in both the major markets under this special safeguard clause.

So, there is no compelling evidence of China receiving any favorable treatment in terms of market access to explain this performance differential. India is faring satisfactorily but the performance of China is absolutely outstanding. Do domestic policies and preconditions play any definite role in explaining this difference in the performance trajectory? We propose to carry out that path analysis in this paper where the internal structure of the domestic policy determines the nature of strategic conducts in the international market that influence the observable performance. The time frame considered here is 1995–2012, a 17-year period spanning over initiation of the WTO, abolition of the MFA and strengthening of market-based competitions. On the basis of trade flow data accessed from ITC database and the official site of the Reserve Bank of India, Sect. 2 presents an analysis of relative performance of India vis-à-vis China in the world market in this post-WTO, post-MFA period. Analysis of growth of overall export of T&G has been supplemented by an exploration of the pattern of revealed comparative advantage in different specific product-line, the nature of market concentration/diversification and the type of price competitiveness enjoyed by these countries. Though the performance of China in terms of her global presence is much more commendable than that of India, at least in the international market there is no compelling reason to believe that China enjoyed any preferential advantage from the importing countries' perspective. The following section (Sect. 3) establishes this point by analyzing the impact of MFA phasing out schedule and special safeguard clause imposed on China on the pattern of India–China competition in the major export markets by identifying the time of structural breaks experienced by them in different product-lines in their major export markets like US and EU in the post-MFA period. Section 4 traces the difference in the specific nature of structural constraints faced by the T&G industry in the respective domestic setting of the two countries and the official intervention policies adopted therein to study the special types of strategic routes adopted by both in locating their international markets. Section 5 identifies the product varieties, production organization along the value-chain and other unique aspects

[1]This is to ensure protection against any surge in textile and clothing imports from China that threatened market disruptions and the orderly development of trade.

related to the pattern of specialization on which the prospect of the industry is contingent even in a post-WTO era. Section 6 concludes the paper by extending an overall assessment.

2 T&G Export: The Relative Performance of India and China

To carry out the analysis we have considered information on trade flows in textile (SITC65) and clothing/garments/apparel (SITC84) following SITC classification. The discussion has been carried out at more disaggregated level (SITC 3-digit level) where textile (65) has further been divided into nine different categories starting from cotton-based textile yarn and artificial yarn, fabrics made of those- both woven and knitted including crocheted, embroidered, other made-up articles inclusive of floor covering[2] and garments (84) into seven different categories divided according to men's and women's apparel, outer garments, under garments, made of man-made as well as artificial fiber, woven or knitted, and so on.[3] Table 1 reports the respective shares of India and China in the total world trade over the selected period (1995–2012) in combined T&G and within the total for each country the relative share of textile.

The simple annual average rate of growth of the value of world trade in T&G is 6.8 % and that for the share of India and China are 2.46 and 11.18 %, respectively. The relative share of textile in this total for each country has been reported in the table and it has been observed that it is more or less stagnant around one-half in India and one-third in China. Thus, the component wise growth profile of textile and garments should not be much different from the overall trends in both India and China. The position of China in the world market was more prominent from the very beginning of our reference period, but it has gone up substantially over the years in comparison to India and Fig. 1 clearly illustrates this temporal pattern.

It is apparent from Fig. 1 that the values of T&G exports from both China and India have increased during the period 1995–2012 and the share of China is increasing much faster than the overall increase in the world trade in T&G and this forceful trend culminated into a dominant position in the world market; while India experienced a modest increase of 1.6 % in her share in total world T&G exports, China achieved more than 26 % increase in her share and accounted for more than

---

[2]Textile yarn (651), cotton fabrics, woven (652), fabrics woven of man-made fibers (653), textile and fabrics woven other than cotton (654), knitted or crocheted fabrics (655), tulle, lace, embroidery, ribbons and others (656), special textile fabrics and related (657), made-up articles (658) and floor covering (659).

[3]Outer garments (men) of textile fabric (842), outer garments (women) of textile fabric (843), under garments of textile fabric (844), outer garments of other articles (845), knitted or crocheted under garments (846), clothing accessories of textile fabrics (847) and articles of apparel and clothing accessories (848).

Table 1 Share of India and China in the world textile and garment exports

SITC (65 + 84)	World '000 USD	India shares (%)		China shares (%)	
Year		T&G (65 + 84)	T/T&G (65)/(65 + 84)	T&G (65 + 84)	T/T&G (65)/(65 + 84)
1995	295120925	2.87	51.38	12.97	36.74
1996	302510093	3.03	53.85	12.39	32.87
1997	327369351	2.93	54.61	14.04	30.52
1998	320468908	2.87	48.73	13.46	30.07
1999	314922249	3.03	49.60	13.77	30.42
2000	336501395	2.93	48.34	15.60	31.09
2001	327201223	2.92	50.11	16.43	31.67
2002	344400178	3.26	49.93	18.05	33.41
2003	391782577	3.44	50.44	20.25	34.25
2004	437996050	3.37	51.68	21.86	35.28
2005	464070283	3.39	48.82	24.95	35.85
2006	509245292	3.26	48.19	28.43	34.01
2007	564806905	3.28	49.25	30.52	32.91
2008	592652838	3.69	48.65	31.52	35.47
2009	511425145	3.63	43.20	32.83	36.04
2010	587203356	3.47	53.38	35.35	37.38
2011	692950790	3.61	51.16	35.99	38.29
2012	655328057	4.14	52.51	39.08	37.62
Average g/r	6.8 %	2.46 %		11.18 %	

Source Authors' estimates based on ITC database

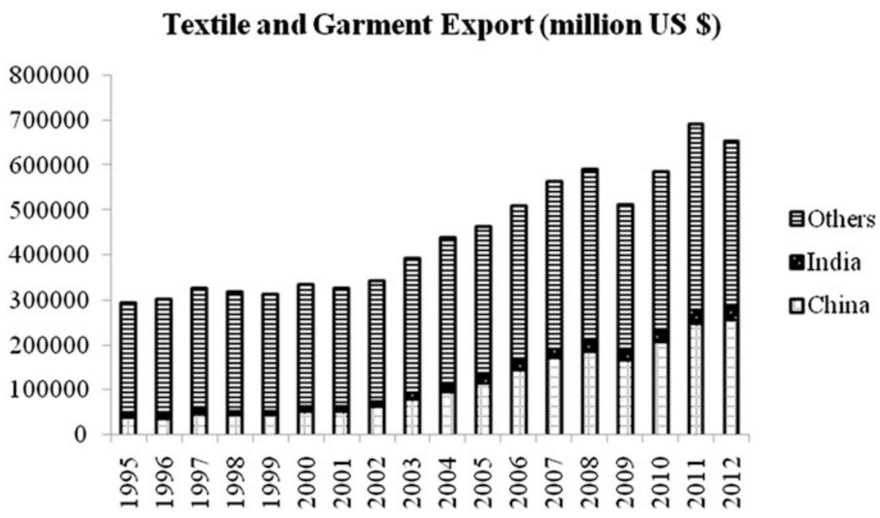


Fig. 1 Position of India and China in T&G export. *Source* Extracted from Table 1

39 % of world's total T&G exports by 2012. At a more disaggregated level it would be interesting to identify the subcategory of products at the three digit level in which China is enjoying comparative advantage vis-à-vis India.

Both India and China revealed comparative advantages in most of the component groups of textile (SITC65) and garments (SITC84) in the world market at the 3-digit level of disaggregation.[4] First, we concentrated on the post-MFA period (2005–2012) where the forces of market signals are expected to play uninterrupted role. China has revealed comparative advantage across all textile and garment groups at the SITC 3-digit level, as RCA indices are greater than one for all these categories. India also experienced RCA value greater than unity in all textile and garment groups except categories 655 (knitted/crocheted fabrics) and 657 (special yarn/fabrics) over the same period. For the convenience of analysis, the commodity groups are classified under 3 heads according to the absolute value of RCA index: [RCA < 1], [1 ≤ RCA < 3] and [RCA ≥ 3] and reported in Table 2. Out of 9 subcategories of textile, China has RCA greater than 3 for 5 cases and India has this very high international presence for 6 cases. For garments, this significant presence is noted for 6 out of 7 subcategories for China and only 3 for India. Thus, India performs relatively better in Textile market and China in Garment market and even in textile the superiority of India lies in the woven products.

It is important to recall here that the RCA index has been constructed in terms of relative importance of the product in the export basket of the concerned country vis-à-vis its world share; however, that does not mean that in absolute scale also these products would enjoy prominence in the export basket of the respective countries. To verify that correspondence, we have identified those product subgroups in T&G for each country, for which the export shares never dropped down below 5 % over the period under consideration. The results reported in Table 3 show that both India and China are specializing according to their revealed comparative advantage as most of the categories identified as important[5] in the export basket of the individual countries also have RCA index values 3 or more.[6] Table 3 reports the situation before and after the withdrawal of the MFA and accordingly two subperiods have been defined as (1995–2004) and (2005–2012).

In terms of the effect of abolition of the MFA in 2005 some marked differences are observed in the reaction pattern of India and China. For China the only category that lost importance in the post-2005 period compared to the pre-2005 period is 848

[4]$\mathrm{RCA}_{ij} = \frac{X_{ij}/X_j}{X_{iW}/X_W}$; i.e., the relative share of commodity i in the export basket of country j has to be compared with the share of ith commodity in the total world trade. If for any country (j) in any product (i) this share exceeds unity, then the country is said to reveal comparative advantage in that particular product.

[5]The important T&G categories are identified as those with share of at least 5 % or more in the total export of T&G of the respective countries for all the years during the period under consideration.

[6]The only exception is category 845 (outer garments and other articles) for India. In category 845 though India has a RCA > 1 for all the years in the post-2005 period, unlike the other important T&G categories, here the RCA value is generally less than 3.

Table 2 Value of RCA index for T&G subgroups of India and China: 2005–2012

SITC	RCA	Country					
		India			China		
		(0, 1)	[1, 3)	3 and more	(0, 1)	[1, 3)	3 and more
	Description						
65	***Textile***						
651	Textile Yarn			√		√	
652	Cotton fabrics woven			√			√
653	Man-made woven fabrics			√			√
654	Woven textile fabrics nes			√		√	
655	Knit/crochet fabrics	√					√
656	Tulle/lace/embr/trim etc.		√				√
657	Special yarn/fabrics	√				√	
658	Made-up textile articles			√			√
659	Floor covering etc.			√		√	
84	***Garments***						
842	Men's/boys' outerwear woven		√				√
843	Women's/girls' clothing woven			√			√
844	Under garments of textile fabric			√			√
845	Outer garments of other articles		√				√
846	Knitted/crocheted under garments			√		√	
847	Clothing accessories of textile fabric		√				√
848	Articles of apparel and clothing accessories other than textile		√				√

Source Authors' estimates based on ITC database

(article of apparel and clothing accessories other than textile). Here, no major shift in commodity composition is noted between the pre and post 2005 period. For India three categories, namely, 652 (cotton fabrics woven), 659 (floor coverings, etc.), and 844 (under garments of textile fabrics) lost their importance in the post-MFA period while category 653 (man-made woven fabrics) and 845 (outer garments of other articles) gained importance in this phase.

The five categories which are important in both China and India's T&G exports are 653, 658, 843, 845, and 846. So, the two countries under consideration are competing in some product groups and specializing in some others. Whether this competition and specialization pattern is grounded on some specific spatial distribution would be interesting to explore. Both India and China are heavily dependent on the European Union (EU) and the North American (NA) markets for their T&G exports (Table 4). These two markets together accounted for nearly 70 % or more of India's total garments exports and 40 % or more of China's total garments

Table 3 Important export categories[a] of T&G subgroups for India and China: 1995–2012

SITC	Export categories	Country			
		India		China	
		1995–2004	2005–2012	1995–2004	2005–2012
	Description				
65	***Textile***				
651	Textile Yarn	√	√		
652	Cotton fabrics woven	√		√	√
653	Man-made woven fabrics		√	√	√
654	Woven textile fabrics nes				
655	Knit/crochet fabrics				
656	Tulle/lace/embr/trim etc.				
657	Special yarn/fabrics				
658	Made-up textile articles	√	√	√	√
659	Floor covering etc.	√			
84	***Garments***				
842	Men's/boys' outerwear woven			√	√
843	Women's/girls' clothing woven	√	√	√	√
844	Under garments of textile fabric		√		
845	Outer garments of other articles	√		√	√
846	Knitted/crocheted under garments	√	√	√	√
847	Clothing accessories of textile fabric				
848	Articles of apparel and clothing accessories other than textile			√	

Source Authors' estimates based on ITC database
[a]Explains at least 5 % of T&G export

exports during the post-MFA regime (2005–2012). For India almost all the important markets[7] are in the developed countries. Saudi Arabia and UAE are the only two countries outside this EU-NA bloc that account for a significant share of her garments exports. While India mostly caters to the niche markets in EU and NA, China has a relatively more diversified market base with a number of East bound developed and emerging economies like Australia, Hong Kong China, Japan, South Korea, Russia, etc., outside the EU-NA bloc accounting for a significant share of her garments exports.

For textiles, on an average, the EU and NA accounted for more than 40 % of India's total exports and more than 20 % of China's total exports for almost all the years under consideration. Unlike garments in case of textiles a number of countries outside EU and NA including a number of developing countries like Bangladesh,

[7]Important markets are identified as those which account for 1 % or more of the country's total textile/garments export for most of the years during 1995–2012.

Table 4 Major markets for garment and textile exports of India and China

Product	Country	Market	2005	2006	2007	2008	2009	2010	2011	2012
Garment SITC84	India	EU	46.99	46.51	48.79	49.13	51.82	47.96	48.75	44.11
		NA	34.73	34.66	31.79	28.28	25.06	27.43	24.43	24.47
		EU+NA	81.72	81.17	80.58	77.41	76.88	75.38	73.19	68.58
		Other major importers	Saudi Arabia, UAE							
	China	EU	29.11	28.62	27.35	24.47	22.59	19.98	21.24	17.77
		NA	22.98	22.85	20.09	18.56	18.21	17.37	18.30	19.91
		EU+NA	52.09	51.47	47.45	43.03	40.80	37.35	39.55	37.68
		Other major importers	Australia, Hong Kong (C), Japan, Korea (S), Russia							
Textile SITC65	India	EU	19.67	24.53	19.01	25.56	26.78	27.09	27.46	22.87
		NA	20.77	19.78	19.62	17.89	20.97	21.63	20.37	19.99
		EU+NA	40.44	44.31	38.63	43.45	47.75	48.72	47.83	42.85
		Other major importers	UAE, Bangladesh, Egypt, Sri Lanka, Turkey							
	China	EU	11.93	12.14	12.84	13.49	13.03	12.96	12.54	11.60
		NA	13.16	12.72	11.97	11.42	11.79	12.10	10.96	11.26
		EU+NA	25.09	24.85	24.81	24.91	24.82	25.06	23.50	22.86
		Other major importers	Bangladesh, Hong Kong (C), India, Japan, Korea (S), Russia, UAE, Vietnam							

Source Authors' estimates based on ITC database

Table 5 Relative import-share of India and China in major common export markets

Product	Market	India (%)			China (%)		
		2005	2010	Δ	2005	2010	Δ
Garment SITC84	EU	4.13	4.95	0.82	21.74	31.87	10.13
	NA	4.31	4.10	−0.21	27.84	41.95	14.11
	UAE	15.38	16.28	0.90	38.45	41.44	2.99
	Australia	2.25	2.25	0.00	74.67	77.37	2.70
	Japan	0.63	0.78	0.15	80.82	82.09	1.27
	Saudi Arb	10.94	9.93	−1.01	58.12	60.25	2.13
Textile SITC65	EU	3.9	4.65	0.75	8.2	14.5	6.30
	NA	8.25	10.18	1.93	24.85	34.98	10.13
	UAE	12.47	13.87	1.40	37.42	35.21	-2.21
	Australia	6.09	6.76	0.67	28.43	39.68	11.25
	Japan	2.60	1.98	−0.62	53.23	59.20	5.97
	Bangladesh	8.41	18.57	10.16	39.46	40.33	0.87

Source Authors' estimates based on ITC database

Egypt, Sri Lanka, Turkey, etc. constitute important export markets for Indian textile products. China too exports her textile products to a number of countries outside EU and NA and here, besides the major buyer of her garment products the additional destinations include countries like India, Bangladesh and Vietnam. So, for both the countries the export markets for textile is more diversified compared to that of the garments.

It would be interesting to study the relative position of India and China as two competing suppliers in the common importing countries, thus identified. Table 5 reports the share of India and China in the import of T&G for all such major countries for 2005 and 2010. In all the common markets China has more dominant share compared to India and in countries like Australia and Japan, China is serving almost the entire market. For India the relative shares did not change to any significant extent for both product groups, with the only exception of Textile in Bangladesh, where it has gone up by more than 10 % point. For China the relative position is generally improving over time with the only exception of UAE in Textile. In case of Garment, China is enjoying an enhancement of her relative export shares by more than 10 % in the markets of both EU and NA; for Textile, such high increase is observed for NA and Australia. In EU and Japan, the export shares of China have gone up significantly (around 6 % point) where India failed to reveal any significant market invasion.

Change in a country's exports might be due to four factors (a) growth of world export; (b) growth of export to relatively faster growing regions; (c) concentration of export in commodities for which demand is growing relatively fast; and (d) more effective competition with other sources of supply. For the first three factors the export growth will be experienced even by maintaining a constant price as here the expansionary effect will come through the shift of the demand curve. However, the fourth factor, if effective, will lead to an increase in the relative market share through price competition. Constant Market Share (CMS) analysis can isolate the contribution of first three factors in the observed change in export between two points in time and the residual effect is claimed as the change due to price competitiveness.[8]

In terms of both product varieties reflected through the RCA values and choice of export destinations discussed above, China is more diversified compared to India. These traits are expected to perform well as non-price competitive strategies. So, a constant market share (CMS) analysis has been reported next to isolate the contribution of price competitiveness in explaining the performance of these countries in T&G sector in the world market. The CMS analysis is carried out for the period 2007–2012. By following the World Bank's TRADECAN 2005 database, the entire world market (where India is an exporter of Apparel/Textile Products) has been divided into eight blocs, namely, North America, Western Europe, Other Industrialized Countries, Developing Africa, Central and South America, Other Developing America, Developing Asia, and Others. The textile and garment groups are identified by following SITC classification at the three digit level. Table 6 reports the CMS decomposition results for the two countries for textiles and garments.

---

[8]The formulae for CMS decomposition may be developed with the help of the following notations:

$V_{..}$: Value of Total apparel/textile export of India in the base year.

V_{ij}: Value of export of the *i*th component of apparel/textile export to the *j*th country by India in the base year;

(The notations with *prime* [′] represent corresponding final year figures).

r: Percentage increase of total World Export of apparel/textile from the base year to the final year;

r_j: Percentage increase of total World Export of apparel/textile to country *j* from the base year to the final year;

r_i: Percentage increase of World Export of *i*th component of apparel/textile export from the base year to the final year;

r_{ij}: Percentage increase of World Export of the *i*th component of apparel/textile export to country *j* from the base year to the final year;

Therefore,

$V_{..} = \sum_j V_{.j} = \sum_i V_{i.} = \sum_i \sum_j V_{ij}; \ldots\ldots$ (A)

The change in export over two different points in time is given by: $V'_{..} - V_{..}$;

At the first level this can be expressed as:

$V'_{..} - V_{..} = rV_{..} + (V'_{..} - V_{..} - rV_{..})$, where the first term in the RHS isolates the effect of change in World export and the second term is the residual effect of competitiveness. At the second level the effect of change in the direction of trade can be isolated by using relation (A) as follows:

$\sum_j V'_{.j} - \sum_j V_{.j} = rV_{..} + \sum_j \left(r_j - r\right) V_{.j} + \sum_j \left(V'_{.j} - V_{.j} - r_j V_{.j}\right)$, where the second term in the RHS represents the effect of change in the value of export due to change in the direction of trade, and the first and third terms carry their usual meaning. Finally, at the third level, by using relation (A), the effect of change in the commodity composition can be isolated as follows:

$$\sum_i \sum_j V'_{ij} - \sum_i \sum_j V_{ij} = rV_{..} + \sum_j \left(r_j - r\right) V_{.j} + \sum_i \sum_j \left(r_{ij} - r_j\right) V_{ij} + \sum_i \sum_j \left(V'_{ij} - V_{ij} - r_{ij} V_{ij}\right); \ldots\ldots\ldots \text{(B)}$$

Here the third term on the RHS is representing the contribution of change in the commodity composition to the change in apparel export, and the other terms carry their usual meaning.

Table 6 CMS analysis of change in textile and garments exports

Change in export (million USD) 2007–2012	Textiles (SITC65)				Garments (SITC84)			
	India		China		India		China	
	Value	%	Value	%	Value	%	Value	%
Change in total export $V'_{..} - V_{..}$	11357.74	100	79644.39	100	7840.10	100	88203.79	100
Increase in world trade $rV_{..}$	3542.365	31.19	20746.38	26.05	2867.00	36.57	33344.97	37.80
Change in market distribution $\sum_j (r_j - r) V_{.j}$	6714.754	59.12	41794.1	52.48	−643.10	−8.20	2776.16	3.15
Change in commodity composition $\sum_i \sum_j (r_{ij} - r_j) V_{ij}$	32.00915	0.28	3300.439	4.14	−409.50	−5.22	7336.34	8.32
Change in competitiveness $\sum_i \sum_j \left(V'_{ij} - V_{ij} - r_{ij} V_{ij}\right)$	1068.613	9.41	13803.46	17.33	6025.71	76.86	44746.31	50.73

Source Authors' estimates based on ITC database

In terms of the value of total change in export of both textile and garment between 2007 and 2012 China has performed much better than India. The value of the change in total export for China between 2007 and 2012 is more than 12 times that of India in textiles and around 7 times more in case of garments. The decomposition analysis for textile suggests that both India and China has been successful in utilizing the buoyant world demand for textile. For both countries more than 25 % of the change in exports is explained by the increase in world trade in textiles. Market diversification is also an important factor as more than 50 % of the change in exports for both the countries is explained by this factor. The third most important factor is change in competitiveness (17.33 % for China and 9.41 % for India).

The decomposition analysis for garments suggest that growth of trade in apparel products is an important factor as more than 35 % of the change in exports for the two countries is accounted for by this factor. Interestingly, for India no gain is observed either due to market diversification or changing commodity composition (a decline of in market diversification by 8.2 % and drop of 5.22 % due to change in commodity composition). China on the other hand experienced some positive gain in both market and commodity diversification (a gain of 3.15 and 8.32 %, respectively). The most important factor influencing the growth of exports for both the countries is price competitiveness and in percentage terms this effect is higher in case of India (76.86 %) as against 50.73 % for China. In fact, for both the countries the performance of T&G sector in the world market is impressive, but the success of China is dazzling compared to the Indian achievement.

3 India and China After Phasing Out of the MFA

In terms of the effect of abolition of the MFA/ATC in 2005 some marked differences are noted in the previous section on the reaction pattern of India and China. For China there was a significant export growth without any major shift in commodity composition between the pre and post 2005 period, however, for India the growth rate is moderate and the composition pattern is less stable. This result is somewhat surprising as China did not get the benefit of quota relaxation for the first two phases and even though the MFA was officially withdrawn on January 1, 2005, the restriction on China was continued in some product ranges till 2008 on the ground of "special safeguard provisions". In fact, it has been noted in Mayer (2005) that the benefits of quota removal will be distributed unequally across countries as quotas are bilateral and their restrictiveness varies across countries. So, China is supposed to face more stringent entry barriers compared to India. However, at the same time Mayer (op. cit.) also pointed out that the significant increase in the market share of China in export of T&G in almost all its constituent categories may be partly due to a streamlining of trading procedures with the removal of the "rules of origin" clause under the MFA/ATC leading to an elimination of the middlemen in Hong Kong China SAR, Macao China SAR and Taiwan Province of China that used to channel apparel produced in Mainland China to the world market. The reported performance of China in pre-2005 era was a definite understatement; in fact, the share of China in the export of the Greater China area has gone up from 45.55 % in 1995 to 95.66 % in 2004 indicating possible channelization of the exports of mainland china through the ports of the other neighboring countries (with whom China has strong production ties) to escape this rule of origin restriction.

An analysis of structural break experienced by India and China in two of their most important common export markets like the US and the EU has been presented here[9] to (a) identify the difference in the timing sequence when the free play of market forces and price competition were actually allowed and (b) to study the differential impact of the withdrawal of the rule of origin restriction on the pattern of trade flows. Since all the quotas were bilaterally imposed, so even within the restricted regime the intensity of restriction was not the same for each country for each product-line. In fact, quota limits under the system of restraints were divided into specific limits and group limits. Group limits placed aggregate limits on a subset of the quota categories. Specific limits on the other hand were imposed on the import of specific goods within each group-based quota category. According to the Office of Textile and Apparel (OTEXA) of the USA, "specific limits" were the most restrictive quotas used under the MFA/ATC. It has been noted that among all the countries exporting T&G products to the US market between 1990 and 2004, China faced the highest level of quantity restrictions where 61 % of her T&G export was constrained by specific limits. In contrast, only 20 % of India's MFA groups were subject to these specific limits. China also had a much higher share of binding quotas compared to India. Quota fill rates provide a useful indication of quota restrictiveness. According to Brambilla et al. (2010) a binding quota has been defined as one in which the fill rate exceeds 85–90 % on an average.

Following this classification it has been shown in Table 7 that out of the ten most important MFA groups in the US market, China had binding average fill rates in nine groups, while for India the corresponding group count is only three. In fact, India enjoyed some specific relaxations also and was not subjected to quota restrictions for some categories like underwear, man-made fiber knit shirts and sweaters, etc., where she faces very strong competition from China in the international market. In the absence of any readily available direct listing of most important MFA categories in the EU market for India and China, eleven important common categories have been taken as those for which the fill rates exceed 90 % for either India or China or both (Jackson 2008). The two countries did not face uniform quota restrictions in different T&G categories in these two regions over this phase of market integration. As the quotas were gradually withdrawn during the ten year period from 1995 to 2005, the integration of different subgroups took place at different points in time. For example in the US market the import of underwear, man-made fiber knit shirts and man-made fiber sweaters from China faced

[9]It has been shown in Naumann (2006) also that the import shares of China in both the EU and the US were always greater than 15 % even under quota regime and for India the corresponding shares lay between 6 and 11 % in the EU and 2–4 % in the US.

Table 7 Structural break for India and China in the US and the EU market in T&G Export

	Description	SITC	Quota fill rate (%)		QLR break point	
			India	China	India	China
US category						
300/301	Cotton yarn	6513	12	52	2007	2007
647/648	MNF trousers[a]	8423	93	99	**2008**	**2003**
340/640	Non-knit shirts	8441	99	99	2001	2007
352/652	Men's underwear[a]	8442	…	85	2007	2006
352/652	Women's underwear	8443	…	85	2003	2009
645/646	MNF sweaters	8451	…	98	2001	2004
638/639	MNF knit Shirts	8452	…	95	2003	2007
347/348	Cotton trousers	8459	99	98	2002	2007
313/314/315	Cotton fabrics		54–75	90+	**2007**	**2003**
	T&G	65 + 84			2003	2006
EU category						
2	Cotton fabrics woven	652	62.37	96.41	2004	2003
4	Men's shirts of textile fabrics (TF)	8441	93.83	96.93	2007	2006
5	Jerseys/pullovers/twinsets/cardigan	8451	91.15	99.15	2007	2007
6	Trousers, breeches etc. of TF	8423	91.41	97.27	2003	2003
7	Blouses of TF	8435	92.34	96.16	2008	2007
8	Other outer garments of TF	8429	99.49	93.61	**2007**	**2003**
9 + 20	Bed linen, table linen, etc.	6584	79.56	92.36	**2007**	**2003**
13	Undergarments, knitted/crocheted[a]	8461	…	98.39	2003	2007
15	Coats and jackets of TF	8431	17.13	96.14	2007	2007
26	Dresses, women's of TF	8433	88.11	90.66	2007	2007
31	Under garments women's of TF[a]	8443	…	98.66	2007	2008
	T&G	65 + 84			**2008**	**2006**

Source Authors' estimates based on ITC database, Brambilla et al. (2010), Jackson (2008)
[a]No restriction for India

restrictions under the MFA till the end of 2005 while India did not face any restriction since 1994. The exact time point from which this relaxation of restrictions would impact the flow of exports of different T&G product groups in the US and the EU market is unlikely to be uniform and difficult to be predicted exogenously. An analysis of the time series data of exports of the different T&C subgroups may provide insight regarding this endogenously determined turning point. Quandt's likelihood ratio statistic (QLR[10]) is computed to estimate this structural

[10]The QLR test consists of calculating Chow breakpoint tests at every observation, while ensuring that subsample points are not too near the end points of the sample. That point is taken as the breakpoint for which the value of the QLR is the highest with strong statistical significance.

break of unknown timing for the top ten MFA categories in the US market and the eleven important common categories in the EU market.

For the total T&G exports in the US market India experienced structural break in the year 2003 and for China it came much later in 2006. For almost all categories where China faced quantity restrictions in the US market, her break points were consistently delayed compared to India indicating the prevalence of more stringent entry barrier for China including the provisions allowed under the safeguard clauses. In contrast, for the EU market this break point for total T&G exports for India was noted in the year 2008 while that for China came 2 years earlier in 2006. This is somewhat unexpected as in 2004 China faced quantity restrictions in 27 T&G subgroups while India faced quantity restrictions in only 16 subgroups. In the constrained categories India showed more than 90 % fill rates for only 6 categories while the corresponding figure for China was 13. For most of the other major T&G categories the break points for both India and China were quite close to each other apparently indicating that neither of the country received any special treatment compared to the other under the MFA quota regime in the European market, which is not consistent with the pattern of quota restrictions discussed earlier. Here comes in the crucial role played by the rules of origin clause laid down in the MFA. From the information on top-10 extra EU importers of T&G in the EU market reported in Jackson (op. cit.) it has been noted that between 1995 and 2005, the import share of China in textile has gone up from 2.4 to 8.3 % and the relative position has also improved from the fourth to the first. For India the share increased moderately (from 3.2 to 4.1 %), but the relative position has gone down from the second to the third. The picture is more dramatic for garments. Here the import-share of China alone has increased from 7.7 to 21.6 % and if the share of China and Hong Kong is considered in conjunction, then the change is less compelling—an increase from 13.8 to 23.9 %. This feature points out a clear tendency towards channelizing export through other ports of greater China (including Hong Kong China, Macao China SAR and Taiwan Province of China) to bypass the country-specific quota limits. For garments, China retained her first position and India's position improved marginally from the fourth to the third, with an increase in import-share from 3.8 to 4.1 %. Thus, imposition and relaxation of quota restrictions could not influence Chinese presence in the EU market to any significant extent. However, this practice did not play strong strategic role in the US market. That is mostly because the geographical location of these respective markets. For the US market, export from the Eastern coast of China like Shanghai involves minimum traveling distance and a definite saving in transportation cost.[11] For the EU market, no such advantage is present. So, the US market experience much direct effect of the withdrawal of the MFA/ATC compared to the EU. Another interesting dimension has been unraveled in Jackson (op. cit.) regarding the nature of restrictions faced by China in the EU market under special safeguards. The data identified a few categories that were not

[11]This location based advantage has been discussed in greater detail in the following section.

facing quotas till 2004, but over 2005–2007 they began to experience fresh quota restrictions, mostly binding, under safeguard clause.

So, the outstanding performance of China in the world market of T&G is not explained by the presence of any favoritism from her trading partners. Rather the position is an outcome of her internal dynamics. In the following section attempt has been made to trace the differences in the path followed in domestic economy, if any, in terms of adoption of different policy initiatives for promoting this sector in India and China.

4 Structure of T&G Industry in India and China

The structure of T&G sector in India is diverse and complex. There is organized sector, decentralized sector and down the line weavers, artisans as well as the farmers. The spectrum of technology is widespread right from handmade to semi-mechanical, mechanical and highly sophisticated computer and microprocessor-based technologies. This sector contributes 14 % of industrial production, 4 % of GDP and 17 % of country's export earnings. It provides direct employment to over 35 million people. The textile value chain extends from raw material, i.e., fibers of finished products, i.e., clothing and made-ups, with spinning, weaving, knitting, and processing coming in between as intermediate processes. The large composite mills produce only 3.6 % output in this sector (NCAER 2009) and it is mostly comprised small scale, nonintegrated enterprises. Among the subsectors the spinning operation is the most consolidated and modernized one with higher level of technical efficiency and India has the second largest yarn spinning capacity in the World next to China (USITC 2001). Operations like dyeing and finishing related to fabric processing are the weakest links in India's textile supply chain (Laila and Asokan 2007). Nearly 65 % of world's installed looms are in India. India is the third largest producer of raw cotton in the world. But the yield of Indian cotton (approx. 300 kg/ha) is very low compared to world average (553 kg/ha), and dismal with respect to some countries like China (1064 kg/ha) and Turkey (1151 kg/ha). Moreover, quality-wise Indian cottons are among the most contaminated in the world (USITC op.cit). This reflects the poor storage facilities and methods of handling cotton not only at the picking stage but also during ginning and pressing. This weakness at the stage of procurement of raw material used to make the end products defective. The R&D in this sector was more focused towards correcting the inherent defects of the cotton output without paying much heed to the requirements of the end products like textile and garments that would ultimately put the raw cotton into use.

Till the 1980s, textile policies in India were in general focused towards protecting employment through the adoption of labor intensive techniques of production. Handlooms and power looms used to enjoy various tax exemptions and became more price competitive compared to the large composite mills. Restrictions like Hank Yarn Obligations (HYO)[12] and others on capacity of mills discouraged large-scale productions. Coupled with this, stricter labor regulations lead to higher costs of production deterring capital investments in the sector. In general, policies have been discriminatory towards the composite mills, with taxes levied at every segment of production, in addition to extremely high income taxes. The labor policy made it infeasible for mills to areas with lower infrastructure costs. They had to bear high burden of utility bills as well as the payment obligation against surplus labor (Ramachandran 2001). The consequence of this lopsided “small is beautiful” type policy orientation got reflected into structural anomalies and concentration of downstream segments of the industry in the decentralized sector and decimation of the organized sector. Decreased productivity along with imposed price restrictions made the sector vulnerable in terms of competitiveness. Cotton Textile order 1948, Textile Policy 1978, Textile Policy 1981 were more or less oriented in similar lines. For the garment industry, perhaps the most draconian of all government policies was reservation of garment manufacture only for small scale units. It has not only prevented expansion, but also impeded technological upgradation of the units. As a result, the garment units could neither attain optimal economies of scale, nor produce international quality garments.

National textile policy 1985 witnessed the first reform of the textile sector in terms of lifting of capacity restrictions, equalization of taxations between mills, power looms and independent processing units. However, years of discriminatory policies and strict regulatory environment against the mill sector resulted in large number of sick textile units in India. With more than 100 textile mills being sick,[13] it seems that the hangover of the regulations remain till very recent times. The push of National Textile Policy 1985 was accelerated in the National Textile Policy 2000 with emphasis on modernization/upgradation of technology. Establishment of apparel parks, National Institute of Fashion Design, use of commercial intelligence and IT in textile industry along with increased focus on hosiery and knitted garments were all in lines with increased competition.

With fillip coming from the reforms along with the recent Textile policies, Indian garments and textile sector has observed more consolidation and integration along with modernization in segments like the garments. It is undergoing considerable growth with modernization in recent times and thus creating newer

[12]The HYO relates to the supply of yarn for the handloom sector, and is exempted from excise duty. The HYO is aimed at guaranteeing an assured supply of cheap and coarse yarn to the handloom sector, so that it can, in turn, churn out “cheaper” fabrics. In reality, however, around 40 % of hank yarn are being consumed by power looms at zero excise duty. The HYO was tantamount to granting a subsidy to the handloom sector on the one hand, and taxing the yarn producers on the other. But the yarn producers business suffered because they were forced to produce a fixed proportion of their yarn of below 40s count, which fetched them a lower margin. More importantly, the obligation prevented the yarn producers from upgrading their product portfolio. This affected subsequent stages like fabric and garments too. The HYO thus, militated against the competitiveness of the textile and clothing sectors.

[13]http://www.thehindubusinessline.com/industry-and-economy/more-than-100-textile-mills-shut-down-in-over-3-years-sharma/article4459332.ece.

opportunities for the segment. The features associated with this growth path are a flexible organization of production that can accommodate small batches of increasingly design intensive and higher value product categories. The demand for goods shifted against standardized products and more in favor of customized goods produced in smaller batches and in multiple styles with greater demand for product variety and flexibility and, hence, giving rise to fragmented markets (Roy 2010). In this buyers' market with relatively labor intensive production segments, dereservation and deregulation could make India increasingly competitive in the international market (Devaraja 2011). Products in this segment being highly customized, it can be expected to add more profit and thus increased prospect. The period of restrictions has crippled Indian textile industry in terms of scale economies, rather focus on labor intensive methods have created economics of scope, mainly in the fashion segment. In fact, there are some advantages of the small size of average Indian firms. They have greater flexibility and can cope with a wide range of production, including very small orders. But the most flexible (and arguably the most productive) firms in the textile and garment sector are the power loom factories. Power looms have very significant advantage of short lead time, which is very important for manufacturers who are supplying to niche markets. Expansion in these lines may help India to retain or enlarge the share even in view of tough competition from China. Recent approval of 40 Textile Parks including SEZs[14] indicates that India has an eye in the fashion segment. Moreover, the reforms in the textiles sector and removal of capacity restrictions on mills indicate India's enthusiasm for scale economies too. After a prolonged phase of restricted regime, direct foreign participation in textile sector has become visible in the India. As an indication of growing interest in the production and marketing of Indian textile and garment, a number of buyers have set up their sourcing/liaison office in India and during the first decade of this millennium the share of FDI in Indian textile sector has gone up from a negligible 0.07 to 0.79 % (Chaudhary 2011). Though these steps signal movement in the correct direction, still for both financial as well as technical support, India is still more dependent on the bureaucratic help from the government rather than on the competitive market participation.

The Chinese textile and garments enterprises are either too large or ultra small. The first phase of economic reforms in China in 1978 witnessed new generations of entrepreneurs: the peasants (Tang 2010). Resources including the raw materials were centrally allocated under a low price regime set by the State under a centrally planned economy. Production tasks were designated by administrative orders and the state owned commercial departments had the monopoly rights of marketing finished products. Inefficiency, outdated technology and high overhead costs ultimately resulted in huge losses by the state owned textile companies amounting to USD 1 billion in 1996.[15] In 1997 China witnessed the first draft policy of restructuring the Chinese textile sector and subsequently the Ninth and Tenth Five

[14] http://shodhganga.inflibnet.ac.in/bitstream/10603/1944/9/09_chapter1.pdf.

[15] http://trade.ec.europa.eu/doclib/docs/2005/december/tradoc_126633.pdf.

Year Plans triggered the structural adjustment of production infrastructure of the sector. The State Planning Commission (currently referred to as NDRC) and State Economic and Trade Commission (SETC) reserved huge amount of fund as loans for establishment of new mills as well as upgradation of the existing ones. Focus was on increasing efficiency.[16]

Emphasizing on large-scale production, the slogan was "grasp the large: let's go off the small".[17] Since 1997 China witnessed closure of hundreds of smaller, inefficient textile units. This was encouraged by the State by allowing them to declare bankruptcy and write off debt. During the same time, the government built a USD 1.5 billion fund to ensure the restructuring: to provide for the loss of 1.5 million jobs and the scrapping of 10 million obsolete spindles, or a quarter of the national total.[18] The painful but necessary reorganization resulted in gradual consolidation and concentration of the industry. The textile industry was also opened to domestic, private and foreign participation, particularly in the garment manufacturing industry. The government has actively encouraged the industry to implement upgrading strategies, particularly investment in capital equipment such as textile machinery and provided tax exemption to State owned enterprises, created textile cities, provided export credit insurance. Huge public investment lead to the emergence of textile cities like *Jiangsu* and *Guangdong* with concentration of large number of textile industrial centers where the availability of the state-of-the-art infrastructure played catalytic role in attracting new companies and investors (Biselli 2009). Garments manufacturing, which is relatively more labor intensive, has been almost entirely deregulated and more than 60 % of Chinese exports of high grade apparel depend on imported materials. The effect is discussed in Yeung and Mok (2004), which has shown that consequently the number of firms decreased from 102,500 in 1995 to 24,500 in 2001, the number of employees decreased from 12.43 million to 5.07 million and the annual gross value of industrial output per firm has gone up 63-fold, turning it into a highly profitable sector by 1999.

Besides domestic investment, foreign direct investment (FDI) has also contributed to the development of China's T&G sector. Hong Kong remains a major investor in China's clothing industry and many Hong Kong-based garment suppliers have their own production facilities on the mainland. The Pearl River Delta (PRD) region has absorbed the majority of foreign direct investment (FDI) from Hong Kong which is in very close vicinity. The majority of textile and garments manufacturing companies in the region developed and managed by Hong Kong entrepreneurs. In fact, China's textile and garments sector is mainly concentrated in Guangdong, Zhejiang, Jiangsu, Fujian, Shandong and the Municipality of Shanghai (Gu 2011). These locations account for nearly 82 % of China's exports of T&G products in 2002. In contrast, clusters of textile and clothing industry in India are highly fragmented and spreading over different regions of the country. Garments

[16] http://trade.ec.europa.eu/doclib/docs/2005/december/tradoc_126633.pdf.

[17] http://trade.ec.europa.eu/doclib/docs/2005/december/tradoc_126633.pdf.

[18] http://trade.ec.europa.eu/doclib/docs/2005/december/tradoc_126633.pdf.

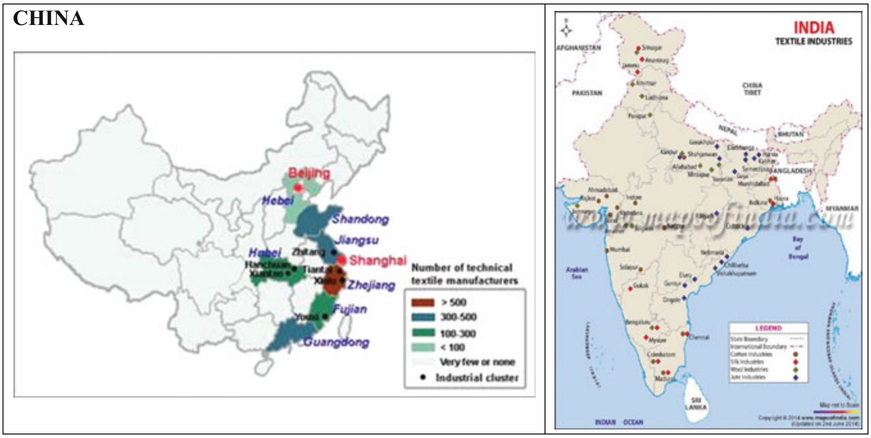


Fig. 2 Regional spread of T&G enterprises in India and China. *Source* fibre2fashion.com. *Source* Maps of India (www.mapsofindia.com)

clusters are found in Okhla (Delhi), Noida, Bangalore, Ahmedabad, Mumbai, Jabalpur, Madurai, the knitted garments clusters are in Tirupur, Ludhiana, Kanpur and the hosiery clusters in Tirupur, Ludhiana (Fig. 2). Indian companies suffer on an average a 37 % cost disadvantage in shipping containers of clothing products from Mumbai/Chennai to the east coast of the US relative to similar container shipments originating from Shanghai (Adhikari and Yamamoto 2007). Moreover, inefficiencies of Indian ports are leading to unusual delays and raising transaction costs (Verma 2002). Koshy (1997) also observed that India's printed fabrics and designs are well known all over the world, but Indian firms are not rated by the importers for their speed of delivery and quality of fabrics.

Compared to China, energy and capital costs are higher in India (USITC op.cit, Adhikary and Weeratunge 2007). The Chinese workers are more efficient and better paid compared to their Indian counterparts and it has been shown in Tewari (2006) that China provides 70 h of training per year to its workers and managers compared to 10 h in India. Average hourly compensation for Chinese garment workers in 2002 was 0.68 USD and the corresponding figure for an average Indian garment worker was 0.38 USD. Chinese workers are more productive than Indian workers also. China concentrates in large-scale production to reap the benefit of scale economy whereas India is in a better position in terms of her English speaking advantage to concentrate on customized products (Tang, op.cit.).

It is apparent from the foregoing discussion that China has utilized the opportunities offered by removal of MFA quotas much better than India. Since China and India are both members of the WTO, in the post-MFA period these two countries enjoy similar challenges and opportunities in the international market. Hence, the huge difference in the performance of the two countries apparently seems to be driven by their domestic policies. The Chinese T&G policy encouraged large-scale production which enabled the China to enjoy economies of scale. The Indian policy on the other hand emphasized on small scale production which led to fragmentation of the industry and here scope economy plays more important role than the scale economy. The consolidation and modernization of the Indian T&G has begun only recently and the bigger factories in India look like lightweight ones compared to the Chinese scale. Revenue of the top textile companies in China is at least tenfold greater than the best performing companies in India. Thus economies of scale in production possibly gave the Chinese industry a cost advantage over her Indian counterpart and enabled China to make aggressive inroads into the world market. While the Indian policy prevented realization of economies of scale in production it encouraged use of labor intensive techniques of production which led to specific skill formation. India's strength lies in small batch orders and customization with higher degree of flexibility. Use of labor intensive techniques is necessary in the high-valued fashion segment since these products require personalized innovation for product differentiation and preservation of product quality. Use of labor intensive techniques also imparts the much needed capacity to adjust to specific requirements in the process of production that is required for these types of goods. Thus, in case of fashion goods economies of scope are far more important than economies of scale. In the Indian T&G industry use of labor intensive techniques and consequent skill formation has helped the industry to enjoy economies of scope (Tang, op.cit.). With this structural characteristic, India is expected to compete much better in the high-valued segment in the export market compared to the low valued segment and the following section presents that analysis by focusing on the export markets in European Union (EU) and North America (NA).

5 Nature of Competition and Specialization in Major Export Markets

Market share in the world exports can be increased either by increasing the volume of export or the unit value of export or both. T&G is a heterogeneous commodity group consisting of both low unit value as well as high unit value products. Products with higher unit value have elastic demand. To evaluate the future prospects of China and India in the world T&G market it is necessary to analyze the relative performance of the two countries in different segments of the T&G value chain. Since, European Union (EU) and North America (NA) together account for a significant share of exports of both these countries we will concentrate on these two markets for this analysis. We'll start from an assessment of the relative market share of India and China in the T&G import of EU and NA in the post-MFA period. Here two specific years, 2005 and 2012, have been selected for this comparative analysis, for which all relevant information was readily available.

From the official website of the Reserve Bank of India (www.rbi.org.in) values of the Unit Value Index (UVI) for different components of Textile sector, viz., (a) textile yarn (SITC651), (b) cotton fabrics woven (SITC652), (c) textile fibers

Table 8 Share of India and China in the total import of T&G in EU and NA

Product group SITC	Unit value index (UVI)			Import-share in EU (%)				Import-share in NA (%)			
	2005	2012	ΔUVI	India		China		India		China	
				2005	2012	2005	2012	2005	2012	2005	2012
651	98	164	66	0.27	0.29	0.26	0.61	0.05	0.08	0.17	0.39
652	103	144	41	0.09	0.08	0.23	0.26	0.07	0.06	0.21	0.27
653	95	114	19	0.06	0.07	0.48	0.73	0.04	0.07	0.38	0.54
654–658	116	182	66	0.59	0.54	1.42	3.01	1.12	1.83	3.72	6.97
659	130	157	27	0.22	0.19	0.12	0.19	0.45	0.48	0.35	0.46
842–848	115	238	123	2.09	2.63	7.41	15.70	2.69	2.76	13.34	25.45
Total	….	….	….	3.32	3.80	9.92	20.50	4.43	5.28	18.17	34.07

Source RBI, COMTRADE

other than cotton (SITC653), (d) made-up articles of textile materials (SITC 654–658) and (e) floor covering (SITC659) have been collected for both the years under consideration, at a constant base of 1999 − 2000 = 100. For all subsectors of Garment only a common UVI is reported against 'articles of apparel and clothing accessories' (SITC 842–848). If the import markets in EU and NA are competitive, India and China are expected to behave as price takers and face same given market prices; hence, the UVI for India has been taken as representation of relevant unit prices for China also (Table 8). Among the different components of export, carpets (floor covering) was the costliest item in 2005 and by 2012, garments became the dearest one. In fact, the change in the UVI is the highest for garments. The data on relative market shares of each one of these categories of T&G are matched with these unit prices for both the importing countries (India and China) in both these markets (EU and NA). It has been observed that the share of T&G imports from India in EU has gone up marginally from 3.32 to 3.80 % between 2005 and 2012 whereas over the same period, in the same sector, the share of import from China has gone up more than twofold, from 9.92 to 20.50 %. The experience for NA is more or less the same. Here India's share in import has marginally increased from 4.43 to 5.28 % after the withdrawal of the MFA and the corresponding increase in the share of China is from 18.17 to 34.07 %. However, among the components of T&G, in both markets and for both countries, garment alone enjoys the dominant share for both the years under consideration (between 50 and 75 %).

How far India and China are competing in same type of product groups in the markets under consideration can be assessed by constructing similarity index of exports developed by Finger and Kreinin (1979). The proposed index of export similarity is defined by the formula:

$$\mathrm{FKI} = S(ab, c) = \left\{ \sum_i \min[X_i(ac), X_i(bc)] \right\} 100$$

which measures the similarity of the export patterns of countries a and b (India and China) to the market c (NA/EU) and $X_i(ac)$ is the share of commodity i (textile/garment) in country a's export to c and $X_i(bc)$ is the corresponding share for country b. If the commodity distribution of a's and b's exports are identical, then for each i, $X_i(ac) = X_i(bc)$, and the index will take on a value of 100. More dissimilar are the intra-group compositions, the value of the index will move towards 0. With this understanding we first checked the value of FKI with India and China as sources, North America (NA) and European Union (EU) as destinations and textile (SITC65) and garment (SITC84) as product groups and found that for all years between 2005 and 2012 the common area of participation never fell below 65 % and never exceeded 85 %. It is only natural to expect that the extent of this similarity in export structure would go down as the exploration would be made at higher level of disaggregation. It has been found in Balasubramanyum and Wei (2005) that for the garment group India has a comparative advantage in the women's clothing of various sorts and men's shirts. India possesses reputation in the production and export of high-value fashion and designer wear clothing.

From the analysis presented in Sect. 4 it has become clear that the Chinese T&G sector concentrates on export-oriented efficiency gain by utilizing the economy of large-scale production where the industry is reaping benefit of foreign direct investment (FDI), state-of-the-art technology, public provision of good quality infrastructure, localization around the coastal belt, domestic availability of high quality cotton and other inputs and last but not the least, large pool of trained and skilled workers. In contrast, the Indian T&G sector is mostly dependent on the scope economies created by highly talented artisans with superb quality craftsmanship, who are dispersed and diversified over a number of small scale enterprises and are more competent in carrying out high-value customized jobs for their niche markets. If this is the correct representation of the structural difference of the industry in China and India, then one can expect China to invade the world market by catering the large generic demand for homogeneous products in each line of production and India to concentrate more on the high-value fashionable alternatives where product heterogeneity in terms of tailor-made uniqueness is the source of rent. India should concentrate more on the upper side of the value chain and china should not reveal any such value-correspondence as she is comfortable with the expected standard in each line of mass production. To verify this surmise we have calculated the Spearman's rank correlation coefficient between the UVI and the import-share for both countries and both years for both the markets (Table 9).

It is apparent from Table 9 that China has never revealed any statistically significant association between the unit value of the product and its dominance in the export basket for any of her major markets before or after the withdrawal of the MFA. In fact, there is absence of relation between these two series indicating Chinese indifference to any specific product group, irrespective of its unit value. On the other hand, for India a remarkable change in the situation is observed between 2005 and 2012. In the earlier year no significant association was observed in the EU market and a relatively weak association prevailed in the NA market. However, in post-MFA, post-WTO period, 2012, the rank correlation coefficients in both EU

Table 9 Rank correlation between import-share and unit-value-index

Country	Year	Market	Spearman's ρ	*p*-value
India	2005	EU	0.49	0.33
		NA	0.77[a]	0.07
	2012	EU	1.00[b]	0.00
		NA	0.86[b]	0.02
China	2005	EU	−0.14	0.79
		NA	0.31	0.54
	2012	EU	0.60	0.21
		NA	0.60	0.21

Source Calculated on the basis of Table 5
[a]Significant at less than 10 % level; [b]Significant at less than 1 % level

and NA are not only quite high in magnitude, are statistically significant at less than 1 % level. This indicates a gradual tendency of India to concentrate in high-value chain.

6 Concluding Observations

In this paper an attempt has been made to explore the structural factors mostly responsible for the spectacular performance of the T&G sector of China in the world export market in comparison to a moderate performance of India over the last two decades, and especially during the post-WTO, post-MFA period. It is found that the growth in the share of Chinese export surpassed that of the world export by registering a more than threefold increase in the value of trade, whereas for India this increase is rather modest. In terms of RCA values also China has revealed comparative advantage in the World market in all nine branches of textile export and seven branches of garment export. For India the niche market is not so diversified, here the high values of RCA (greater than 3) is noted only for three subcategories of textile products and another three subgroups of apparel and clothing. In fact, the combined RCA index value of T&G for India has gone down from 3.58 to 2.56 from 2005 to 2012 and over the same period the corresponding value of China has remained static around 3.18. In terms of market diversification, India mostly focuses on the markets of the European Union (EU) and that of North America (NA) for garments and in case of textile, a few more Middle-Eastern and South-Asian countries are additionally providing the export markets. In contrast, China has more diversified presence in different parts of the world market. Chinese exports are more price competitive for the textile segment and the Indian exports are more price competitive in the garments segment. Chinese exports are more intensive in imported inputs compared to the Indian situation and more integrated with the international order through the production chain. An extensive analysis of the pattern of market participation during and after the MFA phasing out process has been reported for the two countries in their major common export destinations and the relative insensitivity of china in the EU unearth the evasive route of channelizing export through the neighboring countries like Hong Kong China, Macao China SAR and Taiwan Province of China, etc. So, on the whole the Chinese T&G sector is coping much better with the post-MFA situation and reaping benefits of free market trading more intensively compared to India.

To explain this disparate performance in the world market the role of domestic policy was analyzed for both these countries and it was found that while in India the T&G sector mostly comprised small and diversifies enterprises spread over different parts of the country, in China the sector is mostly driven by large industrial estates with a distinct regional concentration along the eastern coast. Though in India foreign direct investment (FDI) has no visible presence in this sector (less than 1 %) in China the investment from the large business houses from Hong Kong China had distinct and decisive role to play in shaping its performance. In the Indian policy framework, given the strong labor laws and social security commitments the large integrated mills were rather discouraged to expand and for the spinning operations, where the large units were technically more efficient, supply obligations like hank yarn obligation (HYO) to the handloom units imposed some additional constraint on the possibility of profitable operation. Consequently there was a development of specific small scale superspecialized artistic talents, more suitable to cater to the customized high-value demand in the fashion market. India concentrated more on the higher side of the value chain in her niche markets in both EU and NA in the post-MFA period by relying more on the economies of scope rather on the economies of scale. This is apparent from her greater competitiveness in the clothing and apparel segment of the industry compared to in the textile segment. While China banked more on *homogeneity* and *scale* India tagged her fortune with *heterogeneity* (product specificity) and *scope*.

The paper will be concluded by indicating a few more important but unattended dimensions that may have serious implication on the future pattern and position of India as well as China in the global market for textile and garments. The major motivating as well as promoting factor for China in attaining fantastic scale economy was rooted in her outward looking reformation strategies, whereby the Chinese firms became integral part of the global value chain in the textile and garment sector. Demand driven value chain has led to the emergence of intermediaries who have thorough knowledge of the production and trading process in this sector, which enabled them to master the technique of delivering large orders to the buyers' specifications without making any compromise with the time and quality components. The T&G MNCs fron Hong Kong have started acting as intermediaries for global buyers leading to the emergence of a triangular production network[19] defined by Gereffi (1999). So, expansion came through the process of global

[19]This implies that the production is done in one country (usually less developed), organized and coordinated by firms in another country (usually middle-income) and sold to a buyer even in another country (usually developed).

integration in the value chain. In contrast, for India the country's strength lie in her vertically integrated supply chain that enables value generation and enhancement within the country (UNDP 2005). Though the Indian market is limited in scale, it is more protected against unpredictable shocks and crisis in the global economy. Here the performance is moderate, though the achievement is stable.

This moderate performance is likely to gain control over the world market in the foreseeable future. Since the Chinese economy is growing much faster compared to that of India, over time the relative importance of a basic manufacturing sector with a strong rural connection would cease to be an important economic sector. Other more sophisticated and advanced sectors will hold the rein of economic growth. Then the Chinese people will also buy textile and clothing related products from the global market and instead of a leading exporter in the world market, China will turn into a major importer. This will definitely pass the baton to India.

Acknowledgments The insightful comments from anonymous referee are gratefully acknowledged.

References

Adhikary R, Weeratunge C (2007) Textile and clothing in South Asia: current status and future potential. South Asia Econ J 8(2):171–203

Adhikari R, Yamamoto Y (2007) The textile and clothing industry: adjusting to the post-quota world, in industrial development for the 21st century: sustainable development perspectives, United Nations, Department of Economic and Social Welfare, United Nations Publications

Balasubramanyum VN, Wei Y (2005) Textiles and clothing exports from India and China: a comparative analysis. J Chinese Econ Bus Stud 3(1):23–37

Biselli M (2009) China's role in the global textile industry, student research projects/outputs no. 039, MBA 2009. China Europe International Business School, HongKong, China

Brambilla I, Khandelwal AK, Schott PK (2010) China's experience under the multi-fiber agreement (MFA) and the agreement on textile and clothing (ATC). In: Feenstra RC, Shang-Jin W (eds) China's growing role in world trade, NBER, pp 345–387

Chaudhury A (2011) Changing structure of Indian textile industry after MFA (Multi Fiber Agreement) phase out: a global perspective. Far East J Psychol Bus 2(2)

Devaraja TS (2011) Indian textile and garment industry: an overview. ICSSR, New Delhi

Finger JM, Kreinin ME (1979) A measure of 'Export Similarity' and its possible uses. Econ J 89 (356):905–912

Gereffi G (1999) International trade and industrial upgrading in the apparel commodity chain. J Int Econ 48:37–70

Gu Q (2011) The Chinese market for clothing. International Trade Centre, Geneva, ITC Technical Paper, Doc. No. SC-11-212E

Jackson K (2008) Unravelling EU quota on textiles and clothing. http://www.etsg.org/ETSG2008/Papers/Jackson.pdf

Koshy DO (1997) Garment exports: winning strategies. Prentice-Hall, India

Laila PI, Asokan T (2007) WTO and Indian textile trade: emerging challenges. J Global Econ 3(2)

Mayer J (2005) Not totally naked: textiles and clothing trade in a quota free environment. Discussion paper, no. 176, UNCTAD

Naumann E (2006) The multifibre agreement-WTO agreement on textiles and clothing. TRALAC working paper, no. 4/2006

NCAER (2009) Assessing the prospects for India's textile and clothing sector. NCAER Study, New Delhi
Ramachandran V (2001) Export competitiveness and the market for textiles: key issues, evidence from firm interviews and policy suggestions. Center for International development, Harvard University, USA
Roy S (2010) Garments industry in India: some reflections on size distribution of firms. Working paper 05/2010, Institute for Human Development, New Delhi, India
Tang X (2010) Comparing technological innovation of textile industries in India and China, International Food Research Institute. www.indiachinainstitute.org/wp-content/uploads/2010/03/Xiaoyang-Tang-Comparison-Textile.pdf. Accessed April–May 2015
Tewari M (2006) Post MFA Adjustments in India's textile and apparel industry: emerging issues and trends. Working paper no. 167, ICRIER, New Delhi
UNDP, Regional Centre in Colombo (2005) International trade in textile and clothing and development policy options: after the full implementation of the WTO agreement on textile and clothing (ATC), policy paper, UNDP, RCC
USITC (2001) India's textile and apparel industry: growth potential and trade and investment opportunities. Staff Research Study, United States International Trade Commission, Washington DC
Verma S (2002) Export competitiveness of Indian textile and garment industry. Working paper no. 94, ICRIER, New Delhi
Yeung G, Mok V (2004) Does WTO accession matter for the Chinese textile and clothing industry? Camb J Econ 28:937–954

Websites Consulted

http://shodhganga.inflibnet.ac.in/bitstream/10603/1944/9/09_chapter1.pdf, A Study of Productivity and Financial Efficiency of Textile Industry of India, accessed on January 10, 2015
http://trade.ec.europa.eu/doclib/docs/2005/december/tradoc_126633, Study on China's Textiles and Clothing Industry and its Market Expansion Strategy, pdf accessed on March 04, 2015
http://www.thehindubusinessline.com/industry-and-economy/more-than-100-textile-mills-shut-down-in-over-3-years-sharma/article4459332.ece, published on 27 February 2013, accessed on March 04, 2015

Impact of Trade Liberalization on Indian Textile Firms: A Panel Analysis

Subhadip Mukherjee and Rupa Chanda

Abstract The textile industry is one of India's major industries in terms of output, investment, and employment. It accounted for 4 % of India's GDP, 14 % of total industrial production, and 11 % of total export earnings in 2012. The industry employs around 45 million people, second only to agriculture. A wide range of textile products are produced and exported from India. Exports of most textile products registered high growth during the 2005–2009 period. The industry and its various subsectors have also experienced a significant reduction in domestic tariff and nontariff barriers over the last decade in addition to liberalization in the global market following the phasing out of the MFA. This chapter examines the impact of India's trade liberalization on firm-level performance in terms of profitability, sales revenues, and imports of raw material for different subsectors of India's textile industry over the 2000–2009 period. It makes use of firm-level panel data from the CMIE−Prowess database to determine this impact. It also analyzes how this impact has been influenced by various firm-specific characteristics. The main finding from this analysis is that there has been an improvement in firm-level profitability and sales and an increase in imported raw materials due to domestic trade liberalization. The analysis also shows that the effect has been stronger through the input sourcing channel, mainly due to the removal of quantitative restrictions on inputs used by the textile industry and that larger firms have been able to gain more from trade liberalization. The analysis and methodology used for the textile industry can be used in similar firm-level studies for other important industries in India.

S. Mukherjee (✉)
Economics Department, IBS Hyderabad, IFHE University, Hyderabad, India
e-mail: subhadip.mukherjee@ibsindia.org

R. Chanda (✉)
Economics and Social Sciences Area, Indian Institute of Management Bangalore, Bangalore, India
e-mail: rupa@iimb.ernet.in

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_11

1 Introduction

India's textile industry is one of the pillars of growth in terms of output, investment, and employment. This industry engages nearly 45 million people, which is the second highest after agriculture. It accounted for 4 % of the country's Gross Domestic Product (GDP), 14 % of its industrial production, and 11 % of total exports earnings in 2012. According to the Planning Commission's Working Group report on India's manufacturing exports, India's textile and clothing exports are projected to reach around US$64.11 billion by the end of March 2017 (Textile Industry Annual Report 2012–2013), clearly indicating the industry's export potential. The industry has also undergone significant trade liberalization in terms of declines in both tariff and nontariff barriers over the 2000–2009 period. The average final goods tariff for the textile industry as a whole has declined from 38.36 % in 2000 to 10.06 % in 2009. In addition, quantitative restrictions on many textile products have also been removed. This chapter examines the effects of this domestic trade liberalization on various firm-level performance indicators for the Indian textile industry.

The plan of the chapter is as follows. Section 2 provides the current scenario for the Indian textile industry. Section 3 provides a brief review of the literature on trade liberalization and manufacturing with specific focus on Indian manufacturing and the textile sector. Section 4 discusses the data sources, the broad methodology and provides a detailed description of the variables used in the model. Section 5 presents the econometric model used for analyzing the impact of trade liberalization on Indian textile firms and also discusses the results. Section 6 concludes the chapter.

It should be noted at the outset that although this is a sector-specific study that aims to highlight the impact of both tariff and nontariff liberalization on an industry which has huge implications for employment creation, poverty alleviation and exports, its approach and findings have a broader relevance to Indian industry at large. The analysis is pertinent to understanding the issues surrounding the relationship between trade liberalization and industrial as well as firm-level performance in India across a wide range of manufacturing sectors and highlights some of the intermediating factors that shape this relationship. This sector-specific study also addresses a major gap in the existing literature on trade policies and industrial performance in India. Although there are studies that discuss the impact of trade liberalization on industry-level performance indicators such as productivity or exports, there are very few firm-level studies of this impact for specific Indian industries and which look at operational performance measures such as revenues and profits. This study thus hopes to contribute to the nascent literature on new-new trade theory in the Indian context and also to provide some insights on how firm-specific attributes can influence the impact of trade liberalization in a developing country like India. Its methodology and approach can be easily applied to other important industries in India.

It is also important to highlight that the focus of this study is on India's trade reforms and their impact on its textile industry. The impact of changes in external demand conditions and of changes in the external policy environment over the study period, i.e., trends in tariffs on textile imports in other countries and in particular, the phasing out of the Multi-Fiber Agreement over the 1995–2005 period, are not the focus of this study.

2 Scenario

According to the Ministry of Textiles, Government of India, this industry can be broadly classified in terms of the following major subsectors and aspects of the industry[1]:

I. The Organized Cotton/Man-Made Fiber Textiles Mill Industry
II. The Man-Made Fiber/Filament Yarn Industry
III. The Wool and Woollen Textiles Industry
IV. The Sericulture and Silk Textiles Industry
V. Readymade Garments
VI. Handlooms, Handicrafts
VII. The Jute and Jute Textiles Industry
VIII. Textile Exports

The trends in India's registered textile industry in terms of the total number of factories, total employment, gross value added, value of output and fixed capital are given in Table 1.

Table 2 represents the CAGR for the total number of factories, total employment and gross value added over the 1981–2006 period. The data clearly indicate stagnant growth in gross value added for the registered textile industry in India. Although, post 1991, there has been positive growth in the total number of factories and employment, the trend has been rather volatile over this period.

Irrespective of the fact that the registered segment of India's textile industry has not improved much over the 1981–2006 period in terms of GVA, total number of factories and total employment, the sector's export performance has been impressive. A wide range of textile products has been produced and exported from India over this period. Table 3 provides the trends for various types of textile exports over the 1993–2008 period.

Table 4, which represents the CAGR for exports of different textile products, clearly indicates that excepting one category (i.e., *Wool Yarn, Fabrics and made-up*) of textile products, all other aforementioned textile items have recorded positive and higher growth during the 2005–2009 period relative to the 1992–2004 period.[2]

[1]http://texmin.nic.in/sites/default/files/ar_12_13_english.pdf (p. no. 3).

[2]The main reason behind this huge decline in the exports of Wool Yarn, Fabrics and Made-up in 2008 is the negative effects of the rise in the Rupee against the US Dollar, which was followed by the decline in global demand for textiles due to the global economic recession in 2008.

Table 1 Textile and garments sector: registered manufacturing (all India) (value in Rs. Lakhs, others in numbers)

Indicators	1980–81	1990–91	1995–96	2000–01	2004–05	2005–06
Number of factories	11,197	10,912	19,838	16,935	16,917	17,459
Total persons engaged	1,376,040	1,150,098	1,855,997	1,619,617	1,714,601	1,878,855
Gross value added (constant prices 1993–94 = 100)	645,676	1,018,459	1,135,826	1,730,374	1,809,421	2,290,536
Value of output (constant prices 1993–94 = 100)	2,379,186	4,203,280	5,005,354	8,962,758	9,914,056	12,030,010
Fixed capital (constant prices 1993–94 = 100)	442,172	952,184	2,541,362	3,749,166	3,541,447	4,133,933

Primary source Computed from ASI, summary results for the Factory Sector, Various years
Secondary source Productivity & Competitiveness of Indian Manufacturing—Textile & Garments Sector National Manufacturing Competitive Council, 2010 (reproduced from the source)

Table 2 Textile and garments: registered manufacturing (all India)

Indicators	Period I: (1980–81 to 1990–91)	Period II: (1990–91 to 2000–01)	Period III: (2000–01 to 2005–06)
	Compound Annual Growth Rate (%)		
Gross value added (at constant prices)	4.66	5.44	5.77
Number of factories (Nos)	−0.25	4.49	0.61
Total persons engaged (Nos)	−1.78	3.48	3.01

Primary source Computed from ASI, summary results for the Factory Sector, Various years
Secondary source Productivity & Competitiveness of Indian Manufacturing—Textile & Garments Sector National Manufacturing Competitive Council, 2010 (reproduced from the source)

Moreover, Table 5 indicates a positive and steady share of India's textile and clothing exports in world textile and clothing exports over the 1996–2007 period. Although the Indian textile industry imports fabrics, it is a net exporting industry.

Though the textile and clothing industry is one of the largest sectors in India, apart from the spinning segment, it is largely dominated by unorganized and fragmented MSME firms. Hence, it has been recognized that there is an urgent need to strengthen this industry by integrating and modernizing firms in the various subsectors and subsegments of this industry, akin to what has happened in the readymade garments segment (Assessing the Prospects for India's Textile and Clothing Sector, NCAER 2009).[3]

[3]National Council of Applied Economic Research.

Table 3 Textile exports (Rs. millions)

Item	1992–93	1995–96	1998–99	2001–02	2004–05	2005–06	2006–07	2007–08
Cotton textiles	40,931	88,222	118,684	146,980	159,244	203,692	248,195	286,064
Manmade textiles	11,414	25,819	30,276	51,912	92,142	90,299	106,841	132,579
Silk	4014	4454	7497	13,635	1819	19,150	19,556	27,446
Wool yarn, fabrics and made-up	1136	2089	3139	2490	3135	3775	3792	3873
Ready made garments	69,307	122,947	183,636	238,776	294,812	381,537	393,429	378,478
Handicrafts	23,578	33,359	49,502	50,514	45,553	58,198	56,976	60,578
Jute	3551	6211	5816	6119	12,412	13,116	11,652	13,614
Coir and coir manufactures	905	2103	3166	2946	4742	5903	7076	6687
Total textiles exports	154,836	285,204	401,716	513,372	613,859	775,670	847,517	923,940
% textile exports to total exports	28.84	26.82	28.74	24.56	16.35	16.99	14.82	15.16
Total exports	536,883	1,063,533	1,397,518	2,090,180	3,753,400	4,564,180	5,717,790	6,801,248

Primary source Annual Report, Ministry of Textiles, Govt. of India, Various Issues
Secondary source Productivity & Competitiveness of Indian Manufacturing—Textile & Garments Sector National Manufacturing Competitive Council, 2010 (reproduced from the source)

Table 4 Textile exports (value in US$ millions)

Item	April–Oct 2005	April–Oct 2006	April–Oct 2007	April–Oct 2008	Compound annual growth (2005–2008) (%)
Cotton textiles	2349.33	2816.98	2816.26	3765.05	17.02
Manmade textiles	1144.66	1328.54	1689.93	2582.28	31.15
Silk	395.71	395.36	384.78	522.92	9.74
Wool yarn, fabrics and made-up	293.9	271.27	377.45	78.16	−35.69
Ready made garments	4286.12	4643.64	4413.89	6812.19	16.7
Handicrafts	755.95	803.8	719.21	857.94	4.31
Coir and coir manufactures	78.6	80.47	92.23	108.95	11.5
Total textiles exports	9304.27	10,340.06	10,493.75	15,270.13	17.96

Primary Source DGCI&S (2009)
Secondary Source Productivity & Competitiveness of Indian Manufacturing—Textile & Garments Sector National Manufacturing Competitive Council, 2010 (reproduced from the source)

Table 5 India's share in world textile and clothing exports (billion US$)

Year	Textiles			Clothing		
	World export	India's export	India's share in world export %	World export	India's export	India's share in world export %
1996	151.06	4.94	3.27	164.14	4.22	2.57
1997	157.73	5.24	3.32	182.28	4.34	2.38
1998	151.31	4.56	3.01	183.33	4.78	2.61
1999	147.92	5.09	3.44	186.03	5.15	2.77
2000	157.46	5.9	3.74	198.94	6.03	3.03
2001	147	5.38	3.66	195.03	5.48	2.81
2002	152.2	6.03	3.96	200.85	6.04	3.01
2003	169.4	6.51	3.84	225.94	6.46	2.86
2004	194.7	6.85	4	258.1	6.62	2.8
2005	203	7.85	3.9	276	8.29	3
2006	217.992	8.837	4.05	309.593	9.465	3.05
2007	238.126	9.446	3.96	345.39	9.655	2.79

Primary source Annual Report 2002–03 Indian Cotton Mills Federation (ICMF) & 2007–08 Confederation of Indian Textile Industry (CITI)
Secondary source Productivity & Competitiveness of Indian Manufacturing—Textile & Garments Sector National Manufacturing Competitive Council, 2010 (reproduced from the source)

In 2007–08, the unorganized power loom segment contributed the largest share in total cloth production (62 %), whereas the organized mill segment accounted for only 3 % (National Manufacturing Competitive Council 2010). According to the

Table 6 Structure of the Indian textile industry

Sector	Units	Volume (2004–05) (million m^2)	Employment (million)
Organized textile mills	1789	1,526	1
Power looms	0.4 million	37,437	4.8
Handloom	3.5 million	5,722	6.5

Primary source Textile Commissioner, Mumbai
Secondary source Productivity & Competitiveness of Indian Manufacturing—Textile & Garments Sector National Manufacturing Competitive Council, 2010 (reproduced from the source)

NCEUS survey of both unorganized and organized textiles units (i.e., *Spinning, Weaving and Finishing and Wearing Apparel*) based on the NSSO (62 round) and the ASI database, the share of value added in output was much higher for the unorganized MSME textile units than for their large organized counterparts.[4] During the 2006–2008 period, for the MSME firms, the share of value added in output was 52.4 %, whereas for large firms, it was only 24 %. Moreover, the total number of employees engaged in the organized segment of the textile industry was 12.58 million in 2005–06, much lower than that in the unorganized MSME segment of the textile industry, which was 16.98 million in 2004–05 (NCAER 2009). This makes evident the significant role of the unorganized MSME segment in generating value added and employment for India's textile industry. The structure of the textile industry in 2004–05 is represented in Table 6.

Given the fragmented and unorganized structure of India's textile industry, there are several challenges, chief among these being: (1) Lack of technological knowledge; (2) Inadequate supply of cheap raw materials; (3) Lack of marketing skills; (4) Lack of accessibility of institutional credit; (5) Lack of product diversification; and (6) Infrastructural bottlenecks.[5, 6]

A survey conducted by CRISIL (2009), identifies some major challenges faced by Indian textile and garments producers, which include: (1) Infrastructural issues, such as shortage of cheap power sources, water, transport vehicles; (2) Manpower issues, namely, the lack of flexibility in labor laws for hiring and contracting, lack of training institutes, shortage of skilled labour; (3) Technology issues due to a highly fragmented spinning and apparel industry, outdated weaving and knitting technology, lack of processing houses, poor quality indigenous dyes (whereas large firms import dyes); (4) Government policies which affect the availability of cheap cotton for the domestic firms (most of the cotton gets exported or is held by middlemen, leading to price increases), varying taxes across states, inadequate refunds and subsidies, high custom duties for equipment; and (5) competition arising from preferential trade agreements with partners such as Sri Lanka and

[4]Report of the Sub Committee of a National Commission for Enterprises in the Unorganized Sector (NCEUS) Task Force on Contribution of the unorganized sector to GDP (2008).

[5]Rangarajan et al. (2007).

[6]Vasa et al. (2014).

ASEAN, which have negatively impacted the Indian textile industry.[7] Studies such as Ananthkrishnan and Chandra-Jain (2005) and RBI (2004) as well as surveys by the EXIM Bank have similarly highlighted the numerous domestic policy constraints and business environment related challenges which have prevented the Indian textile industry from realizing economies of scale and benefiting from the abolition of MFA quotas in major importing markets, under the WTO's Agreement on Textiles and Clothing.

In view of these recognized challenges, the government has introduced various initiatives, such as the Technology Upgradation Fund Scheme (TUFS), Scheme for Integrated Textile Parks (SITP), Initiatives under the National Textile Policy (NTP) 2000, the Technology Mission on Cotton (TMC), etc., to help Indian textile firms in attracting more investments and improve the overall growth prospects of this industry. There are also initiatives to provide financial support to and promote exports from this industry.

3 Literature Review

Various studies have examined the relationship between trade regimes and the performance of the Indian manufacturing sector. These studies have mainly concentrated on the effects of tariffs while a few have also examined the effects of NTB liberalization on the performance of various Indian manufacturing industries as a whole, in terms of indicators such as productivity and employment. Although these studies do not provide any conclusive results, they do indicate that the impact of trade liberalization is context dependent and is influenced by firm, industry, state as well trade regime specific factors, which may explain why no clear results emerge. Some of the most important studies pertaining to trade reforms and certain aspects of performance in Indian manufacturing and more specifically the Indian textile industry are highlighted next.

Several authors have examined the impact of trade liberalization on productivity in Indian manufacturing across various industry groups. For example, Das (2004) finds that only the capital goods sector has experienced an improvement in productivity over the 1980–2000 period, while other manufacturing subsectors, such as intermediate goods and consumer goods sectors have experienced negative productivity growth over the same period. Chand and Sen (2002) further find that the positive effect on productivity was relatively larger in magnitude for the intermediate goods sectors compared to that for the final good sectors. Ghosh (2013) on the other hand finds a negligible effect of tariff liberalization on industry-level productivity in Indian manufacturing and the dependence of this effect on various

[7]Moreover, India has failed to enter into FTA/RTA with countries in the EU or the US, which are the major importers of textiles and garments: Final Report on Enhancing Competitiveness of Indian Manufacturing Industry. Source: Assistance in Policy Making NATIONAL MANUFACTURING COMPETITIVENESS COUNCIL (CRISIL 2009, pp. 27–28).

industry and state level factors such as the interest rate channel, financial conditions, labour market flexibility as well as macroeconomic factors. Some studies also examine the impact of nontariff liberalization on Indian manufacturing performance given the significant liberalization of quantitative restrictions (QRs) in India in the post-2000 period. For example, studies by Sen (2009), Kumar and Mishra (2008), Mehta and Hasan (2012), Hasan et al. (2007), Das (2008) examine the employment effects of tariff and nontariff liberalization on Indian manufacturing. However, no conclusive results emerge across these studies.

There are also firm-level studies on the impact of trade liberalization on Indian manufacturing. Akin to the industry-level studies, these firm-level studies also find divergent effects of trade liberalization on firm-level performance and show that these effects vary across firms, industries, states, and over time. Hasan (2002), Balakrishnan et al. (2006), Topalova and Khandelwal (2011) find a positive effect of tariff liberalization on firm-level productivity, but that these effects remain concentrated in those industries which are technologically more upgraded and which have a higher propensity to import intermediate technologies and other cheap and higher quality inputs. Bas and Berthou (2011) and Ahsan (2013) find that firms that are more creditworthy and which are in industries that are more oriented toward importing foreign technology and different quality intermediate inputs, experience greater improvements in productivity following trade liberalization. Studies such as Balakrishnan et al. (2006), Kato (2009), Loecker et al. (2012), Goldberg et al. (2010a) also find greater benefits of trade liberalization for those Indian manufacturing firms which continuously diversify their production. Goldberg et al. (2010b), Kathuria (2002), Parameswaran (2010) find a positive and significant role of FDI orientation and R&D intensity in realizing the benefits of trade liberalization. There is thus a sizeable body of literature which finds a link between trade liberalization and performance of Indian manufacturing both at the industry and the firm levels. These studies, however, largely focus on performance measures such as productivity, employment, and exports. Literature analyzing other performance indicators such as market penetration or profitability is scant.

There are some studies that examine the link between India's trade liberalization and various performance measures for the Indian textile industry specifically. For example, Virmani et al. (2004) make use of a multi-sector econometric model and time series data for the 1981–2000 period and simulate 838 equations to identify the impact of alternative tariff reduction scenarios on various performance indicators such as output, employment, invested capital stock, exports, and imports for different Indian manufacturing industries, including the textile industry. For the latter industry, they find a significant improvement in all of these aforementioned performance parameters following tariff liberalization. Landes et al. (2005) also find an improvement in the performance of India's Cotton and Textile Industries due to a reduction in import tariffs and other QRs on imports of cheap raw materials and intermediate inputs. Goldar (2005) highlights the removal of QRs on imports of 1429 items in 2000/01, including a large number of textile items and notes this nontariff liberalization significantly increased the imports of intermediate items such as cotton yarn and fabrics, silk yarn and fabrics, man-made filament/spun yarn (including waste), textile yarn and fabrics of material other than cotton, silk, wool, and manmade filament, and made-up textile articles, etc., over the 2002–2004 period.

Another strand of studies focuses on the impact of external trade liberalization, i.e., the phasing out of the MFA under the WTO's Agreement on Textiles and Clothing on the performance of India's textile and clothing industry as well as firms. Ananthkrishnan and Jain-Chandra (2005) use a Computational General Equilibrium model to project the impact of dismantling quotas on Indian textile exports, under different scenarios regarding the degree of liberalization and safeguards available in the global market. They find that while the MFA phaseout presented an opportunity for India to increase its global market share for textiles exports, this was contingent on the continuation of safeguards on China's textile and clothing exports and that once these safeguards were lifted, Indian textile exports would be adversely affected in a quota-free world unless the Indian industry became more price-competitive. At the firm-level, a study by Edwards and Sundaram (2012) finds that the reduction in MFA-related quotas on India's textile exports has been associated with an increase in the sales of quota-restricted textile products relative to non-quota restricted textile products and that the extensive margin has played a greater role in the adjustment of Indian textile and clothing firms. Another firm-level study by Kar and Kar (2015) finds that Indian textile and clothing firms have become more competitive and more export-oriented in the post-MFA regime, but that this has been accompanied by higher industry level concentration at home, possibly in response to the growing cost competition in the global market. The study further concludes that there is a likely adverse impact on employment and wages given growing industry concentration and capital intensity, unless mitigated by related domestic reforms and liberalization of the domestic textile and clothing industry and entry of foreign firms which can benefit labour through increased employment opportunities.

It is worth noting that all the aforementioned studies on the Indian textile and clothing industry highlight the many internal challenges confronting the sector and the significance of these constraints in shaping the industry's performance under a liberalized external or domestic trade regime. Some of the common issues flagged include the fragmented nature of this industry due to the large number of small-scale producers and the reservations for various textile products under the Small Scale Industries (SSI) list during the pre-liberalization years; lack of modernization and technological upgradation by firms; lack of quality intermediates; and the excise duty structure facing the industry.[8] Hence, the literature suggests that both industry and firm-level characteristics are important in shaping the effects of external or internal trade liberalization on the Indian textile industry. As a result, all

[8]For instance, several studies point to the input side challenges facing Indian textile producers prior to liberalization and specifically to the role of imported intermediate inputs in Ramaswamy and Gereffi (2000) and Verma (2002) identify the challenges faced by domestic apparel producers in importing quality fabrics for their production process on account of high import duties and various nontariff barriers on the imports of these items prior to 2001.

these studies note the need to adopt a variety of policies, including cluster development, technology upgradation and modernization, rationalization of excise duties and reduction in input costs to address these internal challenges and to face increased import competition in the home market or increased competition for exports in the global market.

This study builds upon these previous studies and adds value in three respects. First, it examines the effects of both tariff and nontariff liberalization in the Indian textile industry. Second, it examines the impact on less studied firm-level performance measures in this industry in the Indian context, namely profitability, sales, and imported raw materials. Third, it constructs an industry-specific NTB index using detailed product level information on import conditions.[9] The broad objective of this study is to highlight some important channels through which trade liberalization can affect firm-level performance, the relative significance of these different channels and the role of firm heterogeneity in influencing these effects. The theoretical underpinning for this analysis lies in the new-new trade theory literature by Melitz (2003), Bernard et al. (2003), and others which emphasizes the role of intraindustry reallocation of factors following trade liberalization and the heterogeneous impact of trade policies across firms in an industry.

4 Data Sources and Methodology

This study mainly uses balanced firm-level panel data (942 firms *10 years) for the Indian textile industry over the 2000–2009 period. There are 9 subindustry groups in the Indian textile industry. The firm-level information for different variables (for example, sales revenue, profit after tax, total assets, raw material expenditure, and imported raw material expenditure) are extracted from the Prowess database (version 4.12) provided by the CMIE. All tariff-related information is collected from the TRAINS-WITS online database provided by the World Bank. We also measure the NTB data by using the import conditions data from the Director General of Foreign Trade (DGFT) database, and import data, from the Ministry of Commerce and Industry, Department of Commerce, Government of India.

A fixed effect analysis of all the firm-level performance indicators is undertaken to determine the relationship between different trade liberalization indicators (tariff and NTB) and different firm-level performance indicators, after taking into account unobserved firm-level heterogeneity. This approach is also helpful to identify the effects of tariff policy on firm performance across different subsectors of the Indian textile industry.

[9]The methodology for this construction is explained in the next section. It is to be noted that unlike most studies which classify NTBs in terms of high, low, and medium restrictiveness, this study calculates a specific index value for the NTB measure by taking product specific information on NTBs and their import weights, within the industry.

We examine the impact of trade liberalization on three firm-level performance parameters, namely profit after tax, sales revenue, and imported raw materials consumed. All these aforementioned firm-level variables including the firm-size variable (measured by firm's total assets) are measured in Rs. millions and to make this analysis comparable across different textile subindustry groups, we deflate them by using the industry-level WPI deflator using 2004 as the base year. The WPI deflator for various subindustry groups in the Indian textile industry are obtained from Office of the Economic Adviser, Govt. of India, Ministry of Commerce & Industry, Department of Industrial Policy and Promotion (DIPP).

We measure the final goods tariff for the various subindustry groups in the Indian textile industry by using the average MFN applied rates (variable measured in terms of percentage) over the 2000–2009 period. This information has been collected from the WITS-TRAINS database.

This study also estimates nontariff barriers (NTB) affecting the textile industry given the significant reduction that has taken place in QRs in India over the last two decades. Although it is very hard to find a good dataset to measure NTBs, there are some studies (Das 2003; Pandey 1999), which have attempted to measure NTBs for the period 1980–2000, using the import coverage ratio. In these studies, the measurement of NTBs captures the relative restrictiveness of imports for different industries. The import coverage ratios are defined as the percentage of imports within a product category that are affected by NTBs. Their formulation of the NTB coverage ratio was as follows:

Define $w_i = m_i/\sum m_i$ as the import weight, where m_i = imports of the ith commodity where $\sum m_i$ is the total imports.

$$\begin{aligned} \text{Let}\, n_i &= 1 \quad \text{if there are NTB'}s \\ &\quad\; 0 \quad \text{if there are no NTB's.} \end{aligned}$$

Then, the NTB coverage ratio is defined as $\sum n_i\, w_i$. An alternative is to calculate simple averages of the coverage ratios.

The coverage ratio for each input–output sector is calculated according to the following weighting scheme for each 8-digit tariff line and has been assigned a number:

0 % if no NTB applies to the tariff line (i.e., if no licensing is required)
50 % if imports are subject to special import licenses (SIL)
100 % if imports are otherwise restricted or prohibited.

In our study, we use a similar idea but our construction of the NTB variable differs. As the main objective is to examine the impact of a reduction (full and partial) in nontariff barriers on various industries on firm-level performance, instead of constructing an NTB coverage ratio, we have constructed an inverted version of the NTB measure by reversing the weighting scheme for each 8-digit tariff line used by Pandey (1999) and Das (2003).[10] This is mainly done to capture both the effects of partial and full liberalization policies across industries for the period 2000–2009. We use the following weighting scheme for each 8-digit tariff line:

100 % if no NTB applies to the tariff line (i.e., if no licensing is required) ($n_i = 1$)
50 % if imports are restricted by different import licensing policies ($n_i = 0.5$)
0 % if imports are fully prohibited only ($n_i = 0$)

Then, the Industry-level Inverted NTB coverage ratio is defined as,

$$\text{Industry Inverted NTB}_j = \Sigma n_i w_i \tag{1}$$

where j stands for a particular 4-digit Industry and i represents a product line within that particular industry, $w_i = m_i/\sum m_i$ as the import weight, where m_i = imports of the ith 8 digit level commodity where $\sum m_i$ is the total import of the jth industry.

This above scheme enables us to take into account the effects of those imported items (8-digit HS commodities) whose imports are either free or partially free. This is a value addition to other previously constructed NTB measures, which do not take into account the effects of those imported items which are partially restricted.[11]

Based on the above weighting scheme, we have first assigned an appropriate value to each 8-digit product for every year from 2000–2009. We have next calculated each product's import share at the 4-digit industry level for each of the years. Then, we have applied these values to Eq. (1) to get the NTB index for the entire 4-digit industry as classified by the HS system for the 2000–2009 study period. Finally, we have taken a simple average of these inverted NTB values at the 4-digit level in each of the nine industry groups to get the inverted NTB values for the 9 main groups of textile subindustries for this study.[12]

The data for import conditions (import policy) for each 8-digit product for the period 1999–2009 has been obtained from the Director General of Foreign Trade (DGFT), Government of India.[13] The import data for each 2- and 8-digit industry for the period 1999–2009 has been collected from the Ministry of Commerce and Industry, Department of Commerce, Government of India.

[10]The usual NTB index would give 0's for import free products, hence the reverse formulation.

[11]This is due to the fact that, in other previously constructed NTB measures, both prohibited and restricted imported items were considered to be fully protected and was assumed to have no imports happening over the years.

[12]Dressing and dyeing of fur; manufacture of article, manufacture of carpets and rugs, manufacture of cordage, rope, twine and netting, manufacture of knitted and crocheted fabrics, manufacture of made-up textile articles, manufacture of man-made fibers, manufacture of other textiles, manufacture of other textiles n.e.c, Manufacture of wearing apparel, except fur apparel and spinning, weaving, and finishing of textile.

[13]http://www.eximkey.com/Sec/DGFT/ImportPolicy.

5 The Fixed Effect Models: Description, Analysis, and Findings

5.1 Model Specifications

In this study one of the main objectives is to determine the differential effects of trade liberalization on firm performance in the Indian textile industry and how it differs across various subsectors. Hence, a fixed effect approach has been used to determine these effects, after taking into account unobserved firm-level heterogeneity. In the preliminary model we have done simple OLS regressions with the pooled data, with no control variables. Then we have taken a firm fixed effect model which takes into account unobserved firm-level heterogeneity. Finally, in our final fixed effect model, in addition to the firm (c_i) fixed effect, we have included all the year fixed effects (as denoted by λ_t) and the time varying industry effects (as denoted by ω_{jt}), which has enabled these models to control for different unobserved year and time varying industry level (i.e., various subsectors of Indian textile industry) heterogeneity such as productivity shocks. The various versions of the model for the three selected firm-level performance measures (Y_{ijt}), namely, profit after tax (PAT), sales revenue and imported raw materials consumed, are specified in the following Eqs. (2), (3) and (4).

$$Y_{ijt} = \alpha + \tau\,\text{Trade Liberalization Indicator}_{jt} + \beta\,\text{Total Asset}_{ijt} + \varepsilon_{ijt} \tag{2}$$

$$Y_{ijt} = \alpha + \tau\,\text{Trade Liberalization Indicator}_{jt} + \beta\,\text{Total Asset}_{ijt} + c_i + \varepsilon_{ijt} \tag{3}$$

$$Y_{ijt} = \alpha + \tau\,\text{Trade Liberalization Indicator}_{jt} + \beta\,\text{Total Asset}_{ijt} + c_i + \lambda_t + \omega_{jt} + \varepsilon_{ijt} \tag{4}$$

where, i denotes textile firm, j denotes textile subindustry group and t denotes year.

The trade liberalization indicator variable in the above equations includes industry-level lagged final goods tariff and the NTB index. In all versions of the model, the firm-level PAT has been deflated by the industry-level WPI deflator. The selection of the imported raw materials variable is motivated by the fact that there has been a huge decline in industry level tariffs and NTBs for different subindustry groups within the Indian textile industry during the 2000–2009 period which could have led to increased import intensity of textile industry firms. To understand the latter we also regress the reduction in lagged final goods tariff and NTB on firm-level imported raw materials consumed after taking into account the entire firm, year and time varying industry specific unobserved heterogeneity. In our fixed effect models for profit after tax, sales revenue and imported raw material consumed, we have controlled with firm size (total asset used as proxy variable) apart from the main variables of interest (i.e., lagged tariff or NTB Index).[14]

[14]In our models we have also taken firm age and age square as control variables, but as it has remained as a highly insignificant variable, we have finally dropped this variable from our models of analysis. The results with age and age square are given in column 3 in each of the following table, just for illustration.

5.2 Results and Interpretations

This section discusses the results of the fixed effect models for the aforementioned firm-level performance parameters. As discussed earlier, we have analyzed various versions of all these models. The results of these various versions of the models for all textile firms are given in each of the following tables in separate columns. In each table, columns 1–3 represent the results for the OLS regression in the preliminary versions of the models for all textile firms. In column 1 we examine the effects of trade liberalization on firm performance without taking into account other unobserved heterogeneity and without any firm-level controls. In column 2, we include firm size as a control variable and in column 3 we include firm age and its square as controls. Columns 4 and 5 represent the results for the final versions of the fixed effect models for all textile firms. The analyzed results represented in column 4, take into account unobserved firm-specific fixed effects, while the results represented in column 5, additionally take into account unobserved year and time varying industry effects in the model. It should be noted that in each of the fixed effect regressions (columns 4 and 5), the standard errors are clustered at the firm-level.

5.2.1 Firm-Level Profit After Tax

We next discuss the results for the effects of various trade liberalization indicators on the firm-level profit after tax. The effects of a reduction in the lagged final goods tariff and NTB on firm-level profit after tax for all textile firms are given in Tables 7 and 8, respectively.

Table 7 shows that in all specifications reduction in lagged final goods tariff has increased the firm-level profit after tax for Indian textile firms. The coefficient for the lagged final goods tariff in column 1, which represents the result for the OLS regression of the pooled data, clearly shows that a 1 % decrease in the final goods tariff increases the firm-level profit after tax by 0.043 million. Next, when we control for firm size in the regression (i.e., in column 2), though the magnitude of the estimated coefficient of the lagged final goods tariff reduces, it remains highly significant. The estimated coefficient for the lagged final goods tariff in column 2 indicates a 0.024 million increase in firm-level profits for a 1 % decrease in final goods tariff. In the final versions of the model (Columns 4 and 5) where we control for all the unobserved firm, year and time varying industry heterogeneity, respectively, we find a further decline in the magnitude of the estimated coefficients for the lagged final goods tariff, though they remain significant. In Column 5, where we control for all possible sources of heterogeneity, we find that a 1 % reduction in the final goods tariff increases firm-level profits by 0.014 million. Moreover, the estimated coefficients of total assets (which, measures firm-size) in all specifications are positive and highly significant. This provides evidence of a significant role of firm-size in improving firm-level profit after tax. The estimated coefficient of total assets in Column 5 indicates that, a 1 million increase in total assets improves the firm-level profit after tax by 0.011 million.

Table 7 Profit after tax and final goods tariff

Variables	1	2	3	4	5
	OLS			Firm fixed effect	Firm and year and time varying industry fixed effect
	Profit after tax				
Final goods tariff industry wise	−0.043*** (0.009)	−0.024*** (0.004)	−0.025*** (0.005)	−0.017*** (0.006)	−0.014* (0.008)
Total asset		0.006** (0.002)	0.006*** (0.002)	0.011** (0.004)	0.011** (0.004)
Age			−0.006 (0.009)		
Age square			0.000 (0.000)		
Constant	1.166*** (0.289)	0.432*** (0.115)	0.566*** (0.208)	0.177 (0.168)	−0.141 (0.286)
Firm fixed effects				Yes	Yes
Year fixed effects					Yes
Time varying industry effects					Yes
Observations	7395	5438	5438	5438	5438
R-squared	0.0035	0.0156	0.0159	0.0144	0.0398
Number of firms				942	942

Robust standard errors in parentheses
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Thus the results in Table 7 clearly indicate the positive effects of trade liberalization in terms of a reduction in final goods tariff on firm-level profits in the Indian textile industry. Table 8 indicates the positive effects of NTB liberalization on firm-level profits. It shows that in all specifications, the estimated coefficients of the NTB index remain positive and significant. The coefficients of the NTB index in Columns 1 and 2, which represent the results for OLS regression of the pooled data, indicate that a 1 % decrease in the NTB index increases firm-level profit after tax by 0.047 million and 0.029 million, respectively. Interestingly, in the final versions of the model (Columns 4 and 5) where we control for all the unobserved firm, year and time varying industry heterogeneity, respectively, we find a significant improvement in the magnitude of the estimated coefficients of the NTB index. In Column 5, where we control for all possible sources of heterogeneity, we find that a 1 % reduction in NTB increases the firm-level profits for textile firms by 0.10 million. This clearly indicates a significant role of NTB reduction in improving theprofits of Indian Textile firms over the 2000–2009 period.

Table 8 Profit after tax and NTB

Variables	1	2	3	4	5
	OLS			Firm fixed effect	Firm and year and time varying industry fixed effect
	Profit after tax				
NTB index industry wise	0.047** (0.020)	0.029* (0.016)	0.030* (0.016)	0.072** (0.031)	0.100* (0.053)
Total asset		0.006** (0.002)	0.006** (0.002)	0.012*** (0.004)	0.011** (0.004)
Age			−0.002 (0.008)		
Age square			0.000 (0.000)		
Constant	−4.585** (2.017)	−3.067* (1.650)	−3.097* (1.654)	−7.401** (3.120)	−10.720** (5.293)
Firm fixed effects				Yes	Yes
Year fixed effects					Yes
Time varying industry effects					Yes
Observations	7395	5438	5438	5438	5438
R-squared	0.0004	0.0111	0.0112	0.0141	0.0398
Number of firms				942	942

Robust standard errors in parentheses
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

A comparison of the results presented in Tables7 and 8 provides evidence of the relatively higher effects of a reduction in NTB on firm-level profits compared to a reduction in final goods tariff. This could be the result of a removal of the quantitative restrictions on previously prohibited industrial and agricultural items during the study period, which enabled a sharp increase in the variety of imported products in the domestic market. As studies have pointed out and has been highlighted earlier in the review of literature, the latter has not only intensified product market competition but also contributed to firm-level efficiency. We find that this effect has been stronger than that resulting from a reduction in tariffs. This comparison also indicates a relatively more significant role of the input channel (i.e., through a reduction in NTBs) than of the output channel (i.e., through a reduction in the final goods tariff) in improving firm-level profits in the Indian textile industry following trade liberalization.[15]

[15]See, Goldar (2005, p. 10) for further details.

5.2.2 Analysis of Firm-Level Sales Revenue

We next examine the effects of trade liberalization on firm-level market penetration (i.e., measured in terms of firm-level sales revenue) for the Indian textile industry during the study period. Tables 9 and 10 show the effects of reduction in the lagged final goods tariff and the NTB Index, respectively, on firm-level sales.

Similar to the results shown in Table 7, Table 9 shows that in all specifications a reduction in the final goods tariff has positively impacted firm-level sales revenue in the Indian textile industry. The estimated coefficients of the lagged final goods tariff in Columns 1 and 2 (i.e., the preliminary models without controlling any unobserved heterogeneity) clearly indicate that a 1 % reduction in final goods tariff increases firm-level sales revenue by 0.294 and 0.223 million and the coefficients have remained highly significant. In Column 4, where we control for firm specific unobserved heterogeneity, the magnitude of the estimated coefficient of the lagged

Table 9 Sales revenue and final goods tariff

Variables	1	2	3	4	5
	OLS			Firm fixed effect	Firm and year and time varying industry fixed effect
	Sales revenue				
Final goods tariff industry wise	−0.294*** (0.047)	−0.223*** (0.033)	−0.207*** (0.033)	−0.153*** (0.036)	−0.089* (0.050)
Total asset		0.157*** (0.040)	0.156*** (0.040)	0.228*** (0.067)	0.228*** (0.068)
Age			0.089 (0.087)		
Age square			0.000 (0.001)		
Constant	19.349*** (1.478)	14.019*** (1.097)	11.007*** (1.628)	11.307*** (1.590)	10.770*** (2.204)
Firm fixed effects				Yes	Yes
Time fixed effects					Yes
Time varying industry effects					Yes
Observations	7047	5191	5191		5191
R-squared	0.0057	0.1415	0.1521		0.3354
Number of firms			924		

Robust standard errors in parentheses
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table 10 Sales revenue and NTB

Variables	1	2	3	4	6
	OLS			Firm fixed effect	Firm and year and time varying industry fixed effect
	Sales revenue				
NTB index industry wise	−0.280 (0.216)	0.002 (0.127)	−0.055 (0.127)	0.236* (0.144)	0.583* (0.358)
Total asset		0.159*** (0.040)	0.158*** (0.040)	0.237*** (0.066)	0.228*** (0.068)
Age			0.126 (0.088)		
Age square			0.000 (0.001)		
Constant	40.139* (21.585)	8.401 (12.720)	10.988 (12.783)	−15.982 (14.450)	−51.015 (35.375)
Firm fixed effects				Yes	Yes
Time fixed effects					Yes
Time varying industry effects					Yes
Observations	7047	5191	5191	5191	5191
R-squared	0.0005	0.1337	0.1455	0.3019	0.3223
Number of firms				924	924

Robust standard errors in parentheses
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

final goods tariff variable declines, though it remains significant at the same level. This confirms the positive impact of tariff liberalization on firm performance with regard to sales revenues in the Indian textile industry during the 2000–2009 period. In the final model where we include the entire firm, year and time varying industry effects, the estimated coefficient of the lagged final goods tariff indicates that a 1 % decrease in the final goods tariff increases firm-level sales revenue by 0.089 million. Moreover, akin to the firm-level profit after tax, the estimated coefficients of total assets (which, measures firm-size) in all specifications are positive and highly significant. This indicates the significance of firm-size for firm-level sales revenue. The estimated coefficient of total assets in Column 5 indicates that a 1 million increase in total assets improves firm-level sales by 0.228 million. Thus, Table 9 shows that Indian textile firms as a whole have performed well in terms of improvement in their sales revenue following tariff liberalization.

However, unlike firm-level profits, the firm-level sales revenue of Indian textile firms has not been impacted significantly by a decline in NTBs. Table 10 shows that for the preliminary models (Columns 1 and 2) the estimated coefficients of

NTB index are insignificant. This reflects the fact that NTB liberalization has not significantly affected firm-level sales revenues. However, in the final models (Columns 4 and 5), where we are able to control for various sources of unobserved heterogeneity, the estimated coefficients of the NTB index are positive and marginally significant. This shows a weak but positive causal relationship between NTB liberalization and firm performance with regard to sales revenue in the Indian textile industry for the study period. The latter weak causal effect can be explained by the fact that liberalization of NTBs on inputs would only have an indirect impact on a firm's sales revenues through possible improvements in the production process, whereas it would more likely have a direct impact on variables such as import intensity, productivity and output levels.

Thus, from the results presented above we clearly find evidence of an improvement in firm-level performance of the Indian textile industry in terms of profits and to some extent sales following both tariff and non-tariff liberalization. It is also observed that if a firm is bigger in size, then this gain is larger. From the analysis of firm-level profit after tax we find a stronger effect of the input channel compared to the output channel in improving firm-level profits. Other studies have also noted the importance of the input sourcing channel on account of gains arising from lower costs, increased variety and scale of inputs following trade liberalization.[16] Thus, given the identified significance of the import channel on firm performance, we turn next to examining the effects of trade liberalization on firm-level import intensity in the Indian textile industry for our study period.

5.2.3 Firm-Level Imports of Raw Materials

To measure the impact of trade liberalization on Indian textile firms, we examine firm-level imported raw materials consumed as a study variable. We regress trade liberalization indicators (tariff and NTB) on this aforementioned firm-level import intensity variable, to assess if Indian textile firms have been able to increase their sourcing of imported inputs and derive associated gains from the input channel (i.e., scale, variety, costs) following trade liberalization. As mentioned earlier, contrary to the case of final goods tariffs which would have only an indirect impact on firm level import intensity, NTB reduction is expected to have a direct impact due to the removal of import restrictions on a number of previously prohibited imported intermediate items. Tables 11 and 12 show the effects of a reduction in the lagged final goods tariff and NTB Index, respectively, on firm-level imported raw materials consumed during the 2000–2009 period.

As expected, Table 11 clearly indicates that the lagged final goods tariff (i.e., the output channel) does have a very weak causal relationship with imported raw materials consumed by Indian textile firms. Moreover, the results show that this weak link disappears after the inclusion of different firm, year and time varying

[16]Hasan (2002), Balakrishnan et al. (2006), Topalova and Khandelwal (2011).

Table 11 Imported raw material consumed and final goods tariff

Variables	1	2	3	4	5	6
	OLS			Firm fixed effect	Firm and year fixed effect	Firm and year and time varying industry fixed effect
	Imported raw material consumed					
Final goods tariff industry wise	−0.062*** (0.013)	−0.049*** (0.012)	−0.049*** (0.012)	−0.016 (0.016)	0.099 (0.086)	0.000 (0.010)
Total asset		0.007** (0.003)	0.007** (0.003)	0.020** (0.009)	0.019** (0.009)	0.020** (0.009)
Age			−0.017 (0.025)			
Age square			0.000 (0.000)			
Constant	3.561*** (0.417)	2.723*** (0.398)	2.776*** (0.511)	1.598*** (0.504)	−2.946 (3.562)	0.984* (0.595)
Firm fixed effects				Yes	Yes	Yes
Time fixed effects					Yes	Yes
Time varying industry effects						Yes
Observations	2794	2023	2023	2023	2023	2023
R-squared	0.0108	0.0288	0.0352	0.0682	0.0833	0.1514
Number of firms				433	433	433

Robust standard errors in parentheses
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

industry level unobserved heterogeneity in our analysis. Although in Columns 1 and 2 (i.e., the results of the preliminary models), the estimated coefficients of lagged final goods tariff are positive and highly significant, in the final versions of the model (i.e., Columns 4 and 5) the estimated coefficients remain insignificant. This clearly indicates that the reduction in final goods tariff has not had a significant impact on the import behavior of Indian textile firms during the study period. On the contrary, Table 12 shows that a reduction in NTB has led to an increase in the imports of raw materials by Indian textile firms. As expected, although the preliminary models (Columns 1 and 2) show an insignificant effect of NTBs on raw

Table 12 Imported raw material consumed and NTB

Variables	1	2	3	4	5	6
	OLS			Firm fixed effect	Firm and year fixed effect	Firm and year and time varying industry fixed effect
	Imported raw material consumed					
NTB index industry wise	−0.069* (0.040)	0.006 (0.026)	−0.003 (0.026)	0.168* (0.102)	0.182* (0.103)	0.199** (0.103)
Total asset		0.007** (0.003)	0.007** (0.003)	0.021** (0.009)	0.020** (0.009)	0.020** (0.009)
Age			−0.013 (0.025)			
Age square			0.000 (0.000)			
Constant	8.913** (4.057)	0.877 (2.626)	1.823 (2.671)	−15.509 (10.227)	−16.984* (10.356)	−18.443* (10.323)
Firm fixed effects				Yes	Yes	Yes
Time fixed effects					Yes	Yes
Time varying industry effects						Yes
Observations	2794	2023	2023	2023	2023	2023
R-squared	0.0892	0.0189	0.0255	0.0827	0.0967	0.1514
Number of firms				433	433	433

Robust standard errors in parentheses
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

material imports, in the final models (Column 4 and 5), where we include the entire firm, year and time varying industry effects, the estimated coefficients of the NTB index clearly indicate a significant positive effect of NTB liberalization on the imports of raw materials by Indian textile firms. Column 5 indicates that a 1 % reduction in NTB increases the imports of raw materials for textile firms by almost 0.20 million. Thus, the above analysis clearly indicates that although the Indian textile industry is highly export-oriented (and as highlighted earlier is a net exporter), trade liberalization especially through the removal of QRs has increased import intensity and is likely to have had a beneficial impact on firm performance through the input channel.

5.3 Discussion of the Main Findings

There are four key results that emerge from the above regressions. First, for all three firm-level performance indicators, namely profit after tax, sales revenue and imported raw materials consumed that are under focus in this study, there is clearly a positive effect of trade liberalization due to a reduction in tariff and NTBs during the 2000–2009 period. Second, the relative impact of tariff versus non-tariff reduction varies across performance measures. While the positive effect of NTB reduction is relatively greater than that for final goods tariff reduction with regard to firm-level profits, the positive effect of tariff liberalization is relatively greater than that for NTB reduction in case of firm-level market penetration as measured by firms' sales revenue. Third, there is clearly the presence of an input channel through which Indian textile firms appear to have benefited following trade liberalization. Notwithstanding the relatively low import intensity of this industry, Indian textile firms are found to have increased their imports of raw materials due to NTB liberalization and this has in turn positively impacted their performance. Finally, firm-size is important in shaping the effects of trade liberalization on firm-level performance indicators. Larger firms experience larger gains.

How can one explain the results obtained above? While it is not possible to conclusively state the reasons for the above findings without more detailed firm level analysis that incorporates additional firm and industry level characteristics, one can draw inferences regarding the channels through which some of the above results arise based on evidence from earlier studies. The first three points highlighted above indicate the importance of the input channel under both tariff and nontariff liberalization in improving firm performance. This result can be explained by the likely effects of increased access to imported inputs on productivity growth, profitability, sales and output that result following tariff and nontariff liberalization for intermediate inputs. The likely channels for this positive impact include increased scale, increased variety, improved quality, reduced costs of imported intermediate inputs, transfer of technology and R&D spillovers. These sources of gains have been highlighted in the endogenous growth literature by Grossman and Helpman (1991), Ethier (1979, 1982), Rivera-Batiz and Romer (1991) regarding the role of foreign intermediate inputs in enhancing growth and have also been highlighted in several theoretical and empirical studies by Lee (1995), Eaton and Kortum (2001), Goh and Olivier (2002), Xu and Wang (1999) and Alfaro and Hammel (2007) which specify the influence of trade liberalization of intermediate and capital goods sectors on firm level performance.

Our results regarding the importance of input liberalization for Indian textile firms also confirm empirical analysis along similar lines for Indian manufacturing firms as a whole, including studies by Topalova and Kandelwal (2011), Goldberg et al. (2010a, b), Loecker et al. (2012). These latter studies find an increased probability of importing capital goods for the average firm, a positive effect of input tariff cuts on the intensive margin of imports of capital goods, and a positive effect of input-trade liberalization on firms' sales, firm productivity growth as well as firms' ability to introduce new products. Some of these studies also highlight the pro-competitive effects of output tariff cuts, which our study also finds. Our results regarding the positive impact of trade liberalization on firm level sales and PAT are also consistent with the findings of previous studies like Kar and Kar (2015). As highlighted in that study, increased market concentration following trade liberalization could be a possible reason for the findings on improved firm-level performance.

The result regarding the importance of firm size in shaping the benefits of trade liberalization confirms the findings of earlier studies regarding the heterogeneous effect of trade liberalization on firms' technology and product choice. Our results show that larger firms are more likely to have captured the benefits of trade liberalization. One can interpret this result as reflecting the advantages large firms possess over small firms, in terms of technology, scale economies, quality, access to credit, among other factors. Studies on Indian manufacturing across industries have also highlighted the importance of credit conditions, technology, assets, among other firm-level characteristics in influencing firm-level performance and our findings confirm the same.[17]

6 Conclusion

The Indian textile industry is a critical part of India's manufacturing strength. It accounts for a huge amount of output and employment. However, it remains highly fragmented and is driven by the unorganized sector. Since 2001, the textile industry has experienced a significant reduction in tariff and nontariff barriers. The imports of a large number of textile items which were earlier subjected to QRs have been liberalized. The result has been increased export orientation and import intensity and a greater variety of imported inputs being used by firms in the textile industry. This has helped Indian textile firms to improve their efficiency through the input channel, consequently benefiting them through improved profits and sales. Further, we find that firm size plays a significant role in helping Indian textile firms to realize the benefits of trade liberalization.

The analysis carried out for the textile industry is relevant in many ways. It highlights the need for industry-specific studies on trade liberalization and the importance of looking at both tariff and nontariff barriers as the relative impact of these barriers may vary across industries and items. The discussion also highlights the need to look at a wide range of performance indicators as the channels through which trade liberalization affects firms vary depending on the indicator concerned. The analysis also highlights the importance of recognizing the potential gains that

[17]Another study (Mukherjee 2015) by one of the authors of this chapter which makes use of the 3rd and 4th census survey data on MSME registered firms finds that these latter firm level features play an important role, wherein the more technologically advanced, modern, creditworthy firms are able to benefit from input and output tariff and nontariff liberalization.

can arise for firms and industries from the input channel, i.e., due to increased availability of differentiated, better quality, and cheaper inputs in the production process and thus the need to move beyond the traditional focus on product market competition for final goods. It also suggests the need to undertake disaggregated subsectoral analysis within individual industries as the adjustment costs and benefits from trade liberalization are likely to vary across product categories in a given industry. Finally, we can conclude that more of such firm-level industry-specific studies can help in improving our understanding of the various firm and industry-related factors that shape the effects of trade liberalization and in identifying firm and industry-oriented policy measures that can help mitigate the adverse effects and enable realization of the benefits of trade liberalization.

References

Ahsan RN (2013) Input tariffs, speed of contract enforcement, and the productivity of firms in India. J Int Econ 90:181–192

Alfaro L, Hammel E (2007) Capital flows and capital goods. J Int Econ 72(1):128–150

Ananthkrishnan P and Jain-Chandra S (2005) The Impact on India of trade in the textiles and clothing sector. IMF working paper 05/214, Washington, DC

Annual Report (2013) Ministry of textiles, Govt. of India, Various Issues. http://texmin.nic.in/sites/default/files/ar_12_13_english.pdf

Assessing the Prospects for India's Textile and Clothing Sector, NCAER (2009). https://www.dnb.co.in/IndianTelecomIndustry/Textile%20overview.asp

Balakrishnan P, Parameswaran M, Pushpangadan K, Suresh Babu M (2006) Liberalization, market power, and productivity growth in Indian industry. J Policy Reform 9(1):55–73

Bas M, Berthou A (2011) The decision to import capital goods in India: firms' financial factors matter. Centre d'études prospective et d'informations internationals (CEPII), Working paper no 2011–06

Bernard A, Eaton J, Jensen JB, Kortum S (2003) Plants and productivity in international trade. Am Econ Rev 93(4):1268–1290

Chand S, Sen K (2002) Trade liberalization and productivity growth: evidence from Indian manufacturing. Rev Dev Econ 6(1):120–132

Das D (2008) Trade barriers measurement: a methodological overview. Unpublished mimeo. Economics and Research Department, Asian Development Bank, Manila

Das DK (2003) Quantifying trade barriers: has protection declined substantially in Indian manufacturing. Economic and Political Weekly, January 31st Issue

Das DK (2004) Manufacturing productivity under varying trade regimes, 1980–2000. Econ Polit Wkly 423–433

Eaton J, Kortum S (2001) Trade in capital goods. Eur Econ Rev 45(7):1195–1235

Ethier W (1979) Internationally decreasing costs and world trade. J Int Econ 9:1–24

Ethier W (1982) National and international returns to scale in the modern theory of international trade. Am Econ Rev 72:389–405

Edwards L, Sundaram A (2012) Trade liberalization and reallocation of production: an analysis of indian manufacturing. University of Cape Town, School of Economics. Aug 2012

Ghosh S (2013) Do economic reforms matter for manufacturing productivity? Evidence from the Indian experience. Econ Model 31:723–733

Goh A-T, Olivier J (2002) Learning by doing, trade in capital goods and growth. J Int Econ 56 (2):411–444

Grossman G, Helpman E (1991) Innovation and growth in the global economy. MIT, Cambridge, MA
Goldar B (2005) Impact on India of tariff and quantitative restrictions under WTO. Indian Council for Research on International Economic Relations, Working paper no. 172
Goldberg PK, Khandelwal A, Pavcnik N, Topalova P (2010a) Multi-product firms and product turnover in the developing world: evidence from India. Rev Econ Stat 92(4):1042–1049
Goldberg PK, Khandelwal A, Pavcnik N, Topalova P (2010b) Imported intermediate inputs and domestic product growth: evidence from India. Q J Econ 125(4):1727–1767
Hasan R, Mitra D, Ramaswamy KV (2007) Trade reforms, labor regulations, and labor-demand elasticities: empirical evidence from India. Rev Econ Stat 89(3):466–481
Hasan R (2002) The impact of imported and domestic technologies on the productivity of firms: panel data evidence from Indian manufacturing firms. J Dev Econ 69(1):23–49
Kar M, Kar S (2015) Trade policy and inequality: evidence from post-MFA India and some theoretical issues. Working paper no. 007, UNU-WIDER working paper series. Jan 2015
Kato A (2009) Product market competition and productivity in the Indian manufacturing industry. J Dev Stud 45(10):1579–1593
Kathuria V (2002) Liberalization, FDI and Productivity Spillovers- An Analysis of Indian Manufacturing Firms. Oxford Economic Papers 54(4):688–718
Kumar U, Mishra P (2008) Trade liberalization and wage inequality: evidence from India. Rev Dev Econ 12(2):291–311
Landes M, MacDonald SA, Singh SK and Vollrath T (2005) Growth prospects for India's cotton and textile industries. US Department of agriculture, Economic research service, 2005
Lee JW (1995) Capital goods imports and long-run growth. J Dev Econ 48(1):91–110
Loecker JD, Goldberg PK, Khandelwal AK, Pavcnik N (2012) Prices, markups and trade reform. NBER working paper no. 17925
Mukherjee S (2015) Doctoral thesis, impact of trade liberalization on the performance of Indian manufacturing MSMEs, submitted to the Indian Institute of Management Bangalore, India
Mehta A, Hasan R (2012) The effects of trade and services liberalization on wage inequality in India. Int Rev Econ Finance 23:75–90
Melitz M (2003) The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71:1695–1725
National Manufacturing Competitiveness Council (2010) Productivity & competitiveness of Indian manufacturing—textile & garments sector, 2010. http://nmcc.nic.in/pdf/textilesclothing_03july2010.pdf
Parameswaran M (2010) International trade and r&d investment: evidence from manufacturing firms in India. Int J Technol Globalisation 5(1):43–60
Pandey M (1999) NCAER Report on Trade Protection in India. National Council of Applied Economic Research, New Delhi, India
Ramaswamy KV, Gereffi G (2000) India's apparel exports: the challenge of global markets. Developing Econ 38(2):186–210
Rangarajan K, Chaturvedi T, Mishra B, Kishore K, Prakash S (2007) Impact of IPR on SMEs with special reference to textiles and processed food. Foreign Trade Rev 42(2):27–65
Report on Enhancing Competitiveness of Indian Manufacturing Industry: Assistance in Policy Making (2009) National Manufacturing Competitiveness Council, CRISIL. http://nmcc.nic.in/pdf/CRISIL_March2009.pdf
Report of the Sub Committee of a National Commission for Enterprises in the Unorganized Sector (NCEUS) Task Force on Contribution of the Unorganised sector to GDP (2008) Working paper no. 2
Rivera-Batiz L, Romer P (1991) International trade with endogenous technological change. Eur Econ Rev 35(4):971–1001
Reserve Bank of India (2004) Report on currency and finance, 2002–03, Mumbai
Sen K (2009) International trade and manufacturing employment: is india following the footsteps of Asia or Africa? Rev Dev Econ 13(4):765–777

Topalova P, Khandelwal A (2011) Trade liberalization and firm productivity: the case of India. Rev Econ Stat 93(3):995–1009

Vasa J, Aggarwal A, Rebernik M (2014) Problem and prospectus of SME sectors in India & China with special reference to chemical, pharma and textile industries. Int J Sci Eng Res 5(2)

Verma S (2002) Export competitiveness of Indian textile and garment industry. Indian council for research on international economic relations, Working paper 94

Virmani A, Goldar B, Veeramani C, Bhatt V (2004) Impact of tariff reforms on Indian industry: assessment based on a multi-sector econometric model. Indian council for research on international economic relations, Working paper 135, 2004

Xu B, Wang J (1999) Capital goods trade and R&D spillovers in the OECD. Can J Econ 32 (5):1258–1274

Trade, Infrastructure and Income Inequality in Selected Asian Countries: An Empirical Analysis

Ajitava Raychaudhuri and Prabir De

Abstract Theoretical and empirical works related to international trade are confined to models which have labour and capital as explicit factors of production. Although income inequality and trade openness have been given importance in the literature, the role of quantity and quality of infrastructure in international trade has not been investigated extensively in this context. Similarly, growth regressions have highlighted the role of infrastructure and trade openness on economic growth independently, while inclusive growth has not received much attention. This study attempts to unravel the interlinkages and interconnections among infrastructure, trade openness and income inequality, using panel data of 14 Asia-Pacific countries at different levels of development. The empirical exercise clearly reveals influence of trade openness and infrastructure on income inequality but the reverse is not necessarily true. Moreover, country-specific factors turn out to be important determinants of trade openness and income inequality. Further, dynamic panel estimates reveal importance of initial values of both income inequality and trade openness as important determinants in the evolution of these variables, thus supporting the persistence thesis, apart from the positive influence of infrastructure as a determining variable.

1 Introduction

The relation between trade, inequality and poverty within countries is not beyond controversy. Most international trade economists have a perspective of a world in which countries exchange goods, factors and ideas. Free trade in goods leads to equalization of factor prices across countries according to the factor price equal-

A. Raychaudhuri (✉)
Department of Economics, Jadavpur University, Kolkata, India
e-mail: ajitava1@gmail.com

P. De
Research and Information System for Developing Countries (RIS), New Delhi, India
e-mail: prabirde@hotmail.com

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_12

ization theorem. In the traditional literature on neoclassical growth models, capital and labour play the central role as two main factors of production. From the perspective of conventional one sector neoclassical growth theory international linkages do not matter, but from the trade perspective they are the crucial determinants.

Under free trade and competitive conditions, trade promotes growth, and growth reduces poverty (Bhagwati and Srinivasan 2003). In the literature on international trade, the issues of income distribution, growth as well as distortions are more or less discussed in terms of endowments of capital and labour, their growth and their relative prices. Countries that initially had a more regulated trade sector experienced an increase in inequality where trade reform, however, does not appear to have significantly affected changes in income distribution.[1] In the income distribution literature, the functional distribution of the two major factors of production again explains the movement of inequality in income distribution over time.[2] Although infrastructure plays the role of a very important catalyst, it gets virtually no explicit mention in the relevant literature on trade and inclusive growth.[3, 4] The neoclassical as well as new trade theories rely heavily on a two-country, two-factor and two-goods model (the 2 × 2 × 2 model), where infrastructure appears mainly as a complementary factor that facilitates trade through reduction of trade cost (Deardorff 2001; Anderson and van Wincoop 2004).

During the recent decades of globalization, economies in Asia and the Pacific grew rapidly until the ongoing global economic and financial crisis appeared in mid-2007. This acceleration of growth, in which international trade has played an important role, has helped Asian and Pacific countries to make impressive strides in economic development (ESCAP 2009). At the same time, empirical evidence suggests that in some supply-constrained larger economies in Asia and the Pacific, such as China, India and Indonesia, prosperity through trade has been accompanied by rise in inequality.[5] In fact, globalization in Asia and the Pacific has resulted in

---

[1]There is strong literature on trade and income distribution supporting the fact that trade liberalization does not necessarily lead to equality of income in the presence of trade distortion. See, for example Edwards (1997), Slaughter (1997).

[2]Refer, for example Campano and Salvatore (2007), Gourdon et al. (2008).

[3]However, development in endogenous growth theory has introduced the possibility of a productive role of public expenditure on infrastructure with an associated possibility of increasing returns to scale (Barro 1990, 1991).

[4]In economics, there is a need to concentrate on economic infrastructure, which includes services from: (a) public utilities, such as power, telecommunications, piped water supply, sanitation and sewerage, solid waste collection and disposal, and piped gas; (b) public works, such as roads, and major dam and canal works for irrigation and drainage; and (c) other transportation sectors, such as urban and interurban railways, urban transport, ports and waterways, and airports (World Bank 1994). Also refer, Canning (2006), Straub (2008), Roland-Holst (2006), Briceño-Garmendia C et al. (2004).

[5]See, for example Asian Development Bank (2007).

growing social and income inequalities.[6] Thus, free trade is not necessarily a pro-poor growth strategy, where we take either the change in first-order inequality as captured by the Gini coefficient, or the change in the number of people below a prespecified poverty line.

There are compelling reasons why rising levels of inequality can slow down growth prospects when distortions affect different income groups unequally in Asia and the Pacific (e.g. the Lao People's Democratic Republic vs. Thailand). Among other things, high levels of inequality can have adverse consequences for social cohesion, quality of institutions and policies, and infrastructure. In turn, social divisions, low-quality institutions and policies, and a lack of infrastructure can have adverse implications for growth prospects of countries in Asia and the Pacific.[7] Since the increases in income inequality can have important implications for the evolution of economic well-being in the long term, it is important to understand the present symptom of linkage, if any, between inequality, trade and infrastructure in developing Asian and Pacific countries. At the same time, a strong positive influence by infrastructure on economic growth does not necessarily ensure inclusive growth.[8] Figure 1 shows the linkage between infrastructure and poverty reduction. It should be noted that while a reduction in income inequality may reflect a strong poverty reduction programme in a country, this is not a necessary condition. The trade costs throw an additional dimension to this. The link between trade, infrastructure and income inequality as well as poverty reduction can be explained in terms of how far the lack of infrastructure might affect small and unorganized producers in developing countries. This is crucial as trade might promote inequality since a number of small producers may not be able to take part in a liberalized trade regime due to lack of infrastructure. As Fig. 1 shows, roads, electricity and irrigation are key factors that increase agricultural and non-agricultural productivity and employment. If the international trade dimension is taken into account, freight carried by sea and air as well as teledensity can be added to this list of infrastructural facilities. However, this also creates the possibility of unequal access to these facilities, since large-scale farmers and non-farm producers could take much

[6]See, for example ESCAP (2001).

[7]For example, Meschi and Vivarelli (2009) analysed intracountry income differences taking 65 developing countries spanning over 1980–1999. Using a dynamic panel data analysis of Arellano and Bond variety, the authors found that trade really does not significantly influence income inequality within countries, although some other control variables like education (through skill formation) and inflation rate do have significant impact. In this paper, the role of infrastructure is highlighted in a similar methodology of dynamic panel data along with trade variable.

[8]The discussion of inclusive growth can move in several directions, but the literature discusses this either in terms of access and affordability of, and access to, infrastructure by the poor or the effect of infrastructure on income distribution (see for example, Brenneman and Kerf (2002) and Winters et al. (2004)). The concept of inclusive growth is used interchangeably with the concept of pro-poor growth.

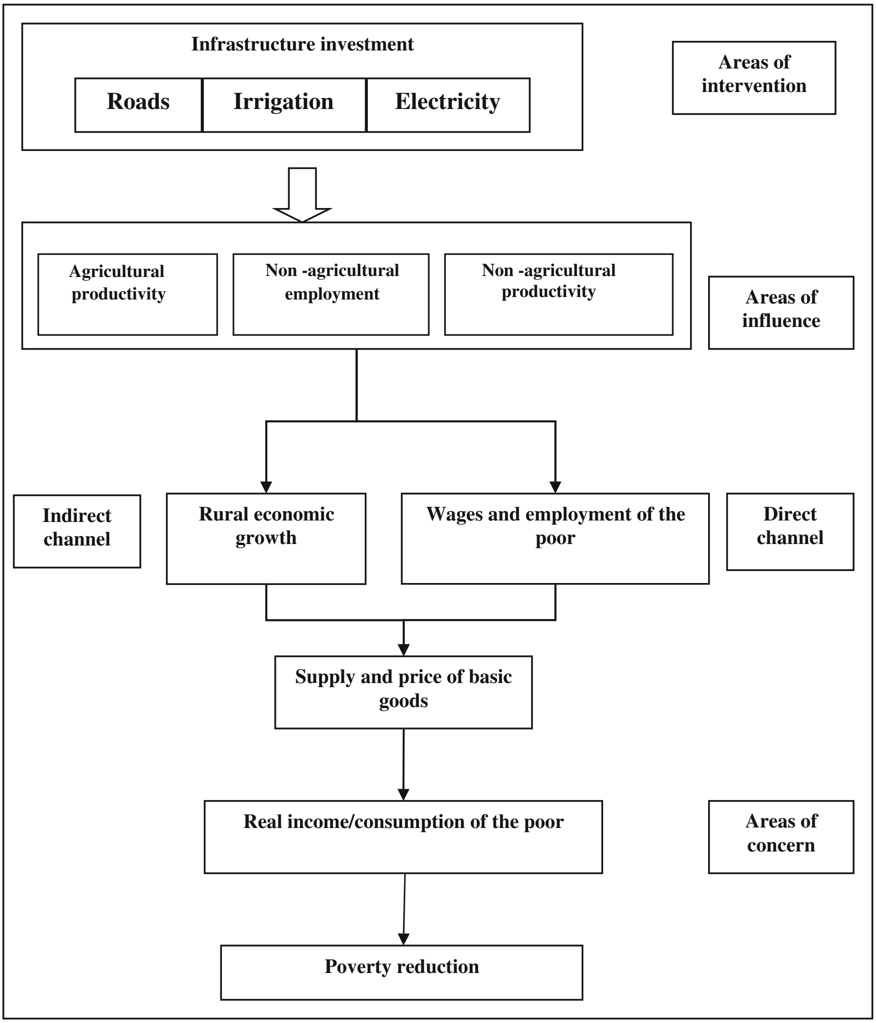


Fig. 1 How infrastructural development helps in poverty reduction. *Source* Ali and Pernia (2003)

greater advantage of these factors than could small-scale farmers and non-farm producers. This paper attempts to provide a succinct answer to the debate on trade, inequality and infrastructure linkages. In addition, it deals explicitly with the recent literature aimed at bringing infrastructure to the forefront in the context of trade and inclusive growth.

Rest part of the paper is organized as follows. Section 2 provides the data and methodology. Section 3 presents stylized facts on infrastructure, trade and inequality in Asian and Pacific countries. The analytical results and possible implications are presented in Sect. 4 and the conclusion is given in Sect. 5.

2 Data and Methodology

To understand the interconnection and interdependence between trade, inequality and infrastructure, cross-country panel data regressions are a common technique used by many researchers, as noted earlier. Hulten (1996) used panel regressions to show the importance of quality of infrastructure to growth, which incidentally has a large number of countries taken from all parts of the world for growth rate in 1990 over 1970. Datt and Ravallion (2002) used an unbalanced panel fixed-effect regression to understand the reduction of poverty in different States in India that was caused by several infrastructural variables. Pinstrup-Andersen and Shimokawa (2006) described a number of studies that used panel data regressions as well as simultaneous regression models to study the impact of infrastructure on agricultural productivity. An excellent analysis comparing different cross-country dynamic panel data estimates, along with more standard pooled OLS and fixed-effect estimates for measuring impact of infrastructure on growth, was carried out by Calderón and Servén (2005).

A good discussion on the technicality of a number of such studies can be found in Levine and Renelt (1991). We have carried out an Asian cross-country study on the same issues, using data from 14 Asian countries: Bangladesh; China; Hong Kong, China; India; Indonesia; Republic of Korea; Malaysia; Nepal; Pakistan; Philippines; Singapore; Sri Lanka; Thailand; and Viet Nam.[9]

While the Appendix provides the definitions of variables and sources, some basics about these variables are presented here. Table 1 shows that inequality (represented by the average Gini coefficient) is quite large at 39.82 and has a range of 23.20–53.10. Trade openness (represented by total trade [export + import] as a percentage of GDP) has an average value of 100.53 % but has a large variation as given by a standard deviation of 104.95. The infrastructure stock has an average of 3.97 with a range of 7.14–2.50. Infrastructure quality (represented by electric power distribution losses as a percentage of output) stands at an average value of 12.93 %, with a range from 2.41 to 35.62 %. The control variables, with the exception of per capita income (PCI), do not have much variation, as shown by their respective standard deviations. There is thus a remarkable disparity between the prosperous and poor Asian countries in terms of per capita income and trade openness.

However, there is a strong correlation between some of these variables (Table 2). Trade openness and per capita income are two variables that are highly correlated between themselves as well as with the population. As the population variable is not correlated with infrastructure stock, they can be used together. Similarly, infrastructure stock is not highly correlated with infrastructure quality, so they can also be paired in a panel regression. Although Levine and Renelt (1991) suggested that collinearity among the independent variables showed a weak dataset

[9]The chosen set contains both LDCs and developing Asian countries since their per capita GDPs—in constant US dollar prices in 2000—range from US$142.39 to US$21,322.16, from 1971 to 2006.

Table 1 Basic statistics

Variable	Obs	Mean	Std. dev.	Min	Max
Inequality (INQ), represented by Gini coefficient	120	39.82	6.69	23.20	53.10
Trade openness (TO), represented by trade–GDP ratio	120	100.53	104.95	9.69	476.78
Infrastructure stock (IS)	120	3.97	1.21	2.50	7.14
Infrastructure quality (IQ), represented by transmission and distribution losses in electricity	120	12.93	7.62	2.41	35.62
Per capita income (PCI)	120	4,010.54	6,893.20	142.39	32,250.00
Geography (Land), represented by per capita availability of land	120	5.62	4.80	0.15	26.19
Inflation (Inf)	120	6.23	5.49	−17.63	24.56
Population (Pop)	120	62.66	6.18	52.45	73.17

Table 2 Correlation matrix

	INQ	TO	IS	IQ	PCI	Land	Inf	Pop
INQ	1							
TO	0.3981*	1						
IS	0.1499	0.0821	1					
IQ	−0.1067	−0.5372*	−0.0935	1				
PCI	0.2670*	0.8778*	0.1316	−0.4877*	1			
Land	0.0473	−0.3533*	0.2216*	−0.0343	−0.4442*	1		
Inf	−0.0900	−0.2794*	−0.2274*	0.2140*	−0.3619*	0.1327	1	
Pop	0.0813	0.6363*	0.1222	−0.6840*	0.6595*	−0.3418*	−0.3368*	1

Note *Significant at the 5 % level

and that "extreme bound" analysis would show how truly independent a variable is, this problem of multicollinearity precludes the use of the full set of regressors.

The following equation was used to explore the relationship of inequality with trade openness and infrastructure:

$$\mathrm{INQ}_{it} = \alpha + \beta_1 \mathrm{TO}_{it} + \beta_2 \mathrm{IS}_{it} + \beta_3 \mathrm{IQ}_{it} + \sum_{l} \delta_l Z_{it} + e_{it} \tag{1}$$

where INQ is inequality measures (represented by the Gini coefficient), TO is trade openness (represented by trade–GDP ratio), IS represents physical infrastructure stock, IQ is infrastructure quality (represented by transmission and distribution loss in electricity), Z is the set of control variables and e is the error term. Here, i represents country and t is the time period. To control for country-level heterogeneity, country dummies were introduced into Eq. (1). The use of country fixed-effects considerably reduces the variance in inequality to be explained and the measurement errors therein.[10] The final estimable equation is

$$\mathrm{INQ}_{it} = \alpha_0 + \alpha_i + \beta_1 \mathrm{TO}_{it} + \beta_2 \mathrm{IS}_{it} + \beta_3 \mathrm{IQ}_{it} + \sum_l \delta_l Z_{it} + e_{it} \quad (2)$$

While Eq. (2) provides the stochastic relationship of inequality with trade openness and infrastructure, it does not indicate the functional relationship of trade openness with inequality and infrastructure. Therefore, to explore the reverse relationship and specifically the relationship of trade with trade openness and infrastructure, the following equation was used:

$$\mathrm{TO}_{it} = \alpha_0 + \alpha_i + \beta_1 \mathrm{INQ}_{it} + \beta_2 \mathrm{IS}_{it} + \beta_3 \mathrm{IQ}_{it} + \sum_l \delta_l Z_{it} + e_{it} \quad (3)$$

An unbalanced panel data has been created from the above. The problem is that the data are driven largely by the availability of information on inequality and infrastructure variables. The difficulty with the inequality data is that although the World Income Inequality Database (WIID) quotes from a variety of sources, there is a major problem with compatibility.[11] The data have been sorted in such a way that they reflect the overall inequality of gross income of a country. An additional problem was the availability of multiple values for 1 year. The WIID database ranks each study in terms of its reliability. The authors were largely guided by that ranking, but if one figure looked like another, it was not used. The data on infrastructure variables also posed a time series continuity problem, which precluded the use of some variables such as paved roads, telephone main lines, railway freight and route (km), education and health-related statistics, etc. As a result, the study could not do proper justice to qualitative variables, as emphasized by Hulten (1996). The number of years in this study is much more but the countries under consideration are much less. Thus, the final estimable model could not accommodate all the variables the authors would have liked to have included. However, authors attempted to utilize the best possible dataset, but the multicollinearity problem precluded simultaneous use of some of the variables even within this reduced set.

[10] In the case of the fixed-effect model, the Least Square Dummy Variable (LSDV) technique was used. The result establishes the fact that those country-specific effects not explicitly mentioned are significant, and clarifies the movement of the explained variable in addition to the common explanatory variables that are taken for each country.

[11] UNU-WIDER World Income Inequality Database, Version 2.0c, May 2008.

2.1 Robustness Checks

The relationships described above cannot be interpreted as causal until the possibility of endogeneity has been ruled out in Eqs. (2) and (3). To address this problem, a dynamic GMM estimator (system GMM)—also known as Arellano–Bover/Blundell–Bond linear dynamic panel data estimation—was used to analyse changes across countries and over time.[12] The estimator also effectively deals with reverse causality by including lagged-dependent variables to account for the persistence of the inequality and/or trade openness indicators.[13]

One of the main advantages of the system GMM estimator is that it does not require any external instruments other than the variables already included in the dataset. It uses lagged levels and differences between two periods as instruments for current values of the endogenous variable, together with external instruments. More importantly, the estimator does not use lagged levels or differences by itself for the estimation, but instead employs them as instruments to explain variations in infrastructure development. This approach ensures that all information will be used efficiently, and that focus is placed on the impact of regressors (such as trade openness) on inequality, and not vice versa.

The reason the cross-country regressions technique was chosen for this study was because it was the methodology used by most of the papers discussed in section B to gain generalizations in the results. Thus, although it is true that few of the country-specific policies and variables (some of them do not vary much over time) may lead to movements in trade and inequality, the cross-country regression generalizes the result, focusing on some important accepted variables that are significant determinants. This in itself is important. Also, the fact can be established that country-specific variables, together with the commonly accepted variables mentioned above, are indeed important catalysts in this analysis.

3 Inequality, Trade and Infrastructure: Stylized Facts

The scatter diagram of inequality with trade openness, per capita income and infrastructure stocks for 14 Asia-Pacific countries in a panel of 1975–2006 shows that inequality is positively associated with all of them (Fig. 2). Trade in countries of Asia and the Pacific as well as rising per capita income have been associated with

[12]First introduced by Arellano and Bond (1991).

[13]Following Arellano and Bover (1995), Blundell and Bond (1998), a system GMM was taken in place of a difference GMM. Arellano and Bover (1995), Blundell and Bond (1998) revealed a potential weakness of the difference GMM estimator. They showed that lagged levels can be poor instruments for first-differenced variables, particularly if the variables are persistent. In their modification of the estimator, Blundell–Bond suggested the inclusion of lagged levels along lagged differences. In contrast to the original difference GMM, they termed this the expanded estimator system GMM.

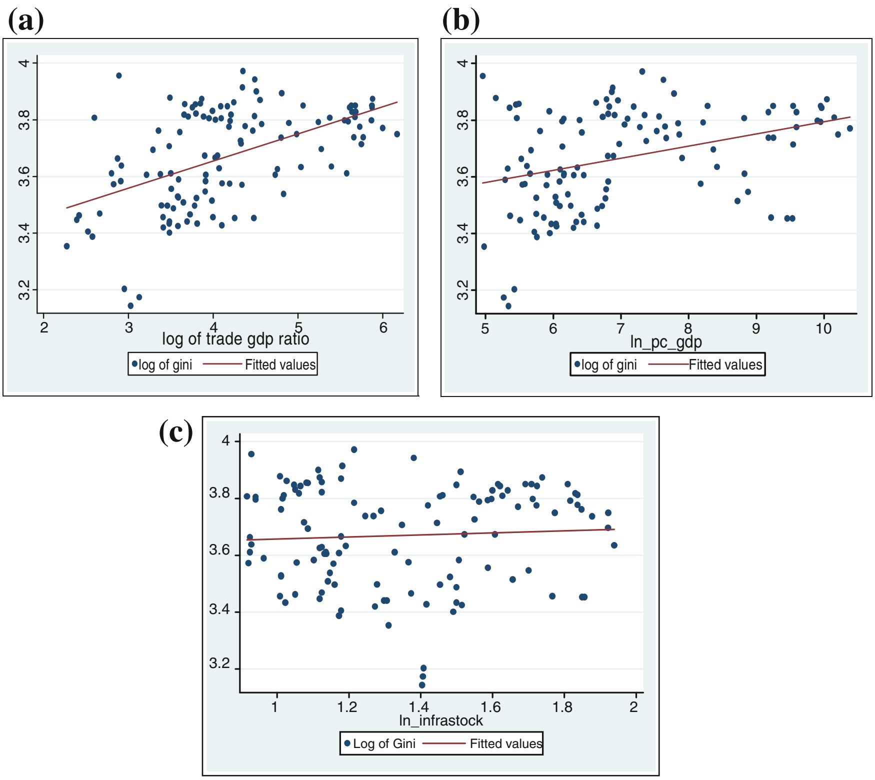


Fig. 2 **a** Inequality and trade openness. **b** Inequality and per capita income. **c** Inequality and infrastructure

higher inequality. Figure 2c suggests that infrastructure stock in Asian and Pacific countries generated inequality, albeit at a lower magnitude, compared with trade openness. Trade openness and infrastructure stock, however, show a positive correlation, suggesting that infrastructure underpins trade openness (Fig. 3), whereas trade openness in the region is driven by trade liberalization with high tariffs in the initial years (Fig. 4). Therefore, it can be concluded that given all other things are constant, Asian and Pacific countries have experienced an increase in inequality. Three important observations in this discussion follow.

First, the larger countries of Asia and the Pacific have witnessed an absolute rise in inequality over time (Fig. 5). China, for example, witnessed a steep rise in inequality, increased from 28.60 in 1975 to 41.50 in 2006 with a low of 23.20 in 1982. Bangladesh, on the other, witnessed a fall in inequality from 45.00 in 1976 to 31.00 in 2006, with a rise of inequality at 43 % in 1997. Thus, cross-country variability in the inequality is very important when we attempt to understand its consequences on trade and infrastructure.

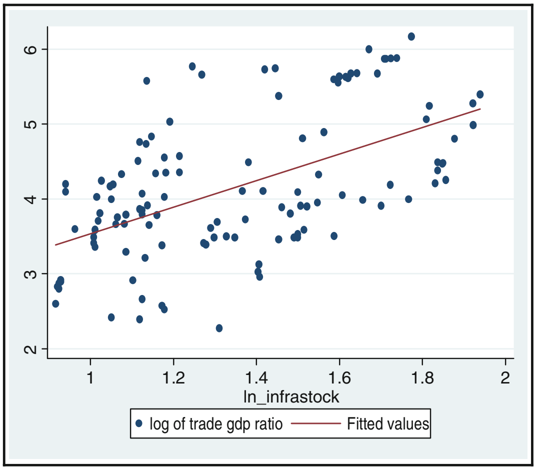


Fig. 3 Infrastructure and trade

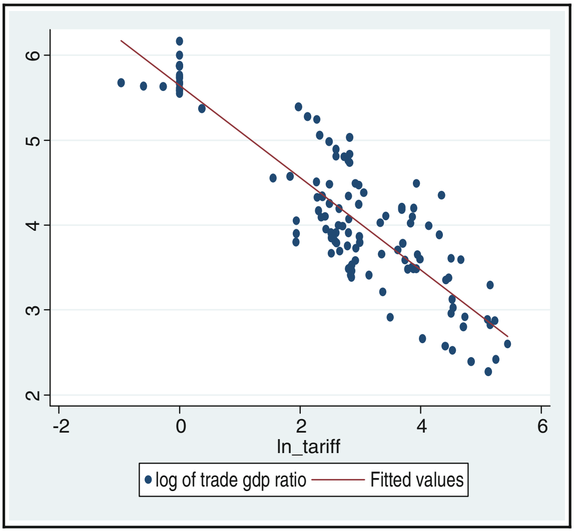


Fig. 4 Trade openness and trade liberalization openness

Second, Asian and Pacific countries have witnessed a rise in trade liberalization (a fall in import tariffs) and trade openness (a rise in trade–GDP ratios), but the trends vary across countries (Fig. 6). For example, the Republic of Korea, Malaysia and the Philippines have witnessed parallel movements of trade liberalization and trade openness from the initial years, whereas others have witnessed a mixed result showing relatively higher distortions in the initial years. Those countries that adopted an export-led growth strategy have also been associated with drastic trade liberalization in terms of tariff cuts. Nonetheless, trade liberalization and trade

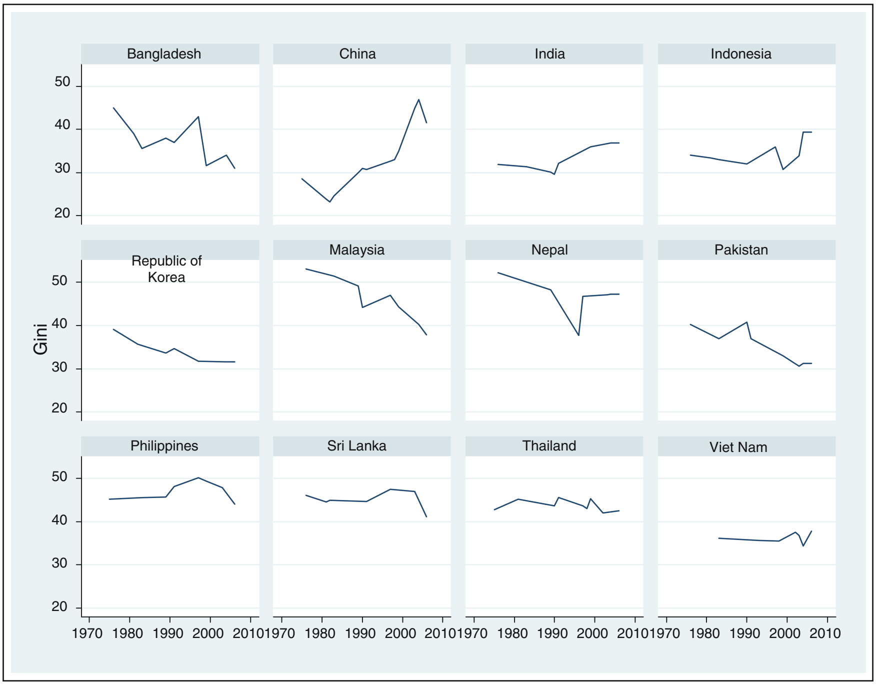


Fig. 5 Inequality trends

openness are uncontested and documented in the development of Asian and Pacific countries.

Third, infrastructure disparity is rampant across countries in Asia and the Pacific, and the stock of infrastructure in those countries is not too large (Fig. 7). For example, China, the Republic of Korea, Indonesia and Malaysia have higher infrastructure stocks than others. Starting with a low base, Indonesia appears to have progressed much faster than the other countries in raising its infrastructure stocks. Therefore, change in infrastructure stock is one of the important factors affecting the long-term behaviour of the Asian and Pacific countries.

4 Trade, Infrastructure and Inequality: The Empirical Relationship

This section begins by exploring whether or not trade openness reduces inequality with reference to Eq. (2). Since the data show nonlinearity, all regressions are run in double logarithmic form. The estimated results are presented in Table 3. The following observations are worth noting.

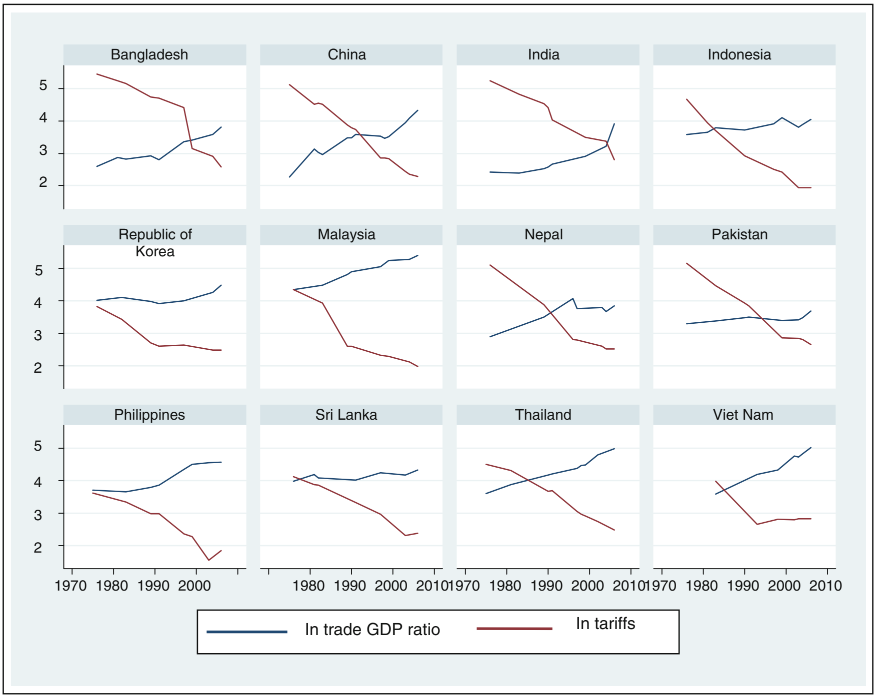


Fig. 6 Trends in trade openness and trade liberalization

There has been a very strong correlation between per capita income and the openness index as revealed by the trade ratio to GDP. Thus, it will create a definite problem if both of these variables are taken together. However, even if we do not take per capita income, it might well influence the result through a fixed-effect regression, where country size and strength are important determinants. This is also important for a random-effect regression, but there are factors that may be beyond the control of individual countries and may thus affect its performance; for example, a global macroeconomic scenario or policies taken in large countries. A similar situation holds for the population variable taken here (see Appendix). Hence, the regressions reported here try to avoid the obvious multicollinearity problem. Also, the data structure shows nonlinearity so that double log regressions give better results than non-transformed variable-based regressions.

The other important point to note is that in all regressions the classical linear regression is dominated by either the random-effect or fixed-effect model, as revealed by the Breusch–Pagan Lagrange Multiplier test. Hence, the ordinary regression results are not reported here because they are not statistically tenable. However, finally it is reported whether a random-effect or a fixed-effect regression model holds as revealed by the Hausman test.

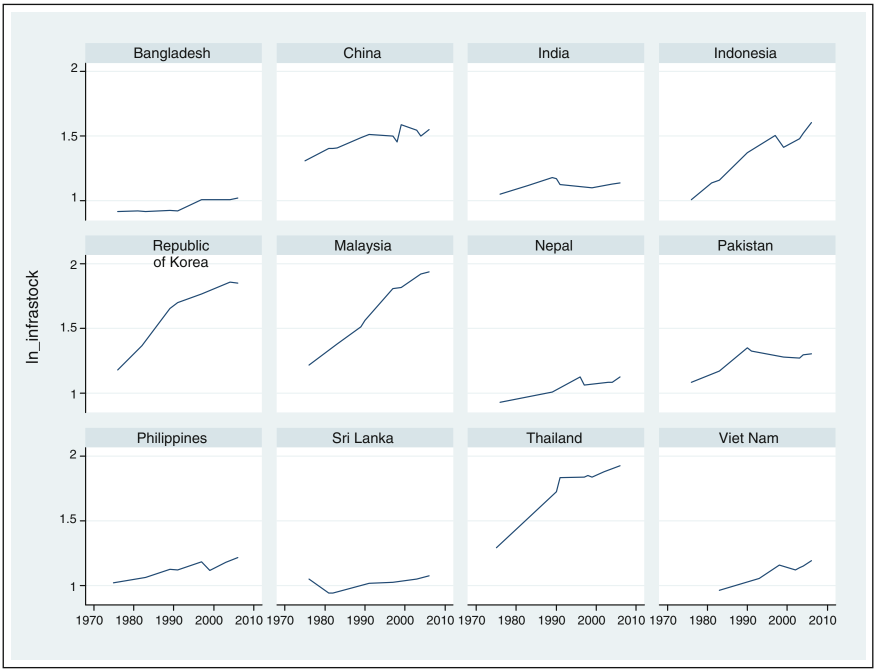


Fig. 7 Trends in infrastructure

Next, the authors have attempted to test the persistence of either inequality or openness among the countries in the panel through a dynamic panel analysis. This system dynamic panel data analysis, following Arellano–Bover/Bover–Bond methodology, also solves the endogeneity problem that is a major headache in any dynamic panel analysis. The system dynamic panel approach improves upon the difference dynamic panel analysis by Arrellano and Bond by reducing the finite sample bias, thereby increasing the efficiency of the estimators. The authors have attempted to maintain some parity between the static panel analysis and the dynamic panel analysis, so that one may judge the persistence theory in the light of an ordinary static panel outcome.

First, the Breusch–Pagan statistic clearly establishes that panel regressions are better than the corresponding pooled classical regressions. Only regression in column 3 of Table 3 shows that the GLS-based random-effect model dominates the fixed-effect model. In the other cases, the fixed-effect model is the relevant one.

Second, as explained above, trade openness and per capita GDP could not be taken together for multicollinearity. In the regression given by column 2, trade GDP in logarithmic terms has a significant positive effect on inequality. Thus, more openness leads to higher inequality, which supports the theory of gains for skilled labour after countries become more integrated with the global economy. A better quality of infrastructure with lower T&D loss reduces inequality, although

Table 3 Inequality, trade and infrastructure: baseline regression results (dependent variable = ln Gini)

Variables	ln Gini	ln Gini	In Gini
(1)	(2)	(3)	(4)
Trade openness (ln trade GDP)	0.145[c] (0.042)	0.107[b] (0.028)	
Infrastructure stock (ln infrastock)	0.020 (0.100)	−0.126[a] (0.076)	−0.038[c] (0.091)
Infrastructure quality (ln tdloss)	0.086[a] (0.034)	0.055[a] (0.032)	0.059[b] (0.027)
Per capita income (ln pc GDP)			0.278[c] (0.032)
Land (ln land)	0.228[b] (0.095)		0.317[c] (0.068)
Constant		3.265[c] (0.169)	
Observations	120	120	120
Adjusted R-squared	0.659	0.331 (non-adjusted)	0.784
Country fixed-effects	Yes	No	Yes
Model	OLS	GLS	OLS
Breusch and Pagan LM test, chi2 (*p*-value)	59.46[c] (0.000)	66.70[c] (0.000)	142.44[c] (0.000)
Hausman test, chi2 (*p*-value)	7.91[c] (0.095)	2.15 (0.542)	12.00[b] (0.017)
Fixed-effect coefficients	Max = 3.229[c] (Indonesia)		Max = 2.186[c] (Singapore)
	Min = 2.284[c] (Singapore)		Min = 1.210[b] (Malaysia)

Note Standard errors in parentheses
[a]$p < 0.1$, [b]$p < 0.05$, [c]$p < 0.01$

geographically larger countries lead to higher inequality. Infrastructure does not have any significant effect on inequality in Asian and Pacific countries. The individual effects were also calculated; they more or less indicate that countries have individual effects (maximum and minimum values are given) that show considerable positive impact on inequality even in the absence of other variables in the regression.

Third, column 3 shows similar results as in the first regression, except that this regression shows that (a) countries do not have any specific individual effect, and (b) better infrastructure reduces inequality. Here, the constant value again plays an important positive and significant role. However, in this case, global policies play a more important role than country-specific policies. The R^2 value is much lower in this case than in the other cases. As expected, the per capita GDP has a similar impact as that in trade openness and, since they move together, inequality is certainly exacerbated as countries have more openness.

Fourth, infrastructure stock is significant in two of the three models. The estimated coefficients of infrastructure stock clearly indicate that infrastructure improvement reduces income inequality, all other things being equal. Controlling for country fixed-effects, 1 % improvement in infrastructure would lead to a fall of

4–12.6 % in inequality for Asian and Pacific countries. A rise in infrastructure stock helps countries to increase income in the medium to long term, thus generating scope for redistribution of income, both within and across countries.

At the same time, the quality of infrastructure is very important. All the estimated coefficients of the variable infrastructure quality are significant at the 5–10 % levels. The estimated coefficients of infrastructure quality (represented here by transmission and distribution loss of electricity) indicate that quality deterioration of infrastructure facilities (e.g. a rise in transmission and distribution loss of electricity) would lead to a rise in income inequality. It is intuitive that higher losses in transmission and distribution of electricity lead to less availability of electricity for users, causing a disadvantage for the poorer income groups. This may well lead to inequality in income distribution as the latter is disadvantaged with regard to one crucial input in the production of goods and services.

Fifth, the estimates also suggest that inequality is high in land-endowed countries (per capita land density) and in countries with higher per capita income in the Asia-Pacific region.

Next, the question of whether or not there is a reverse relationship from income inequality to trade openness is explored with reference to Eq. (3). Here, trade openness is taken as a dependent variable, with income inequality, infrastructure stocks and infrastructure quality, and other control variables taken as independent variables. The estimated results are presented in Table 4, columns (2)–(5). The following observations are worth noting.

First, all regression models again show the dominance of panel regression over classical pooled regression. In addition, all the panel regressions show strong country-specific individual effects. Trade liberalization as measured by tariff levels strongly influences trade openness in regressions expressed in columns 2 and 5. In fact, the first regression shows a very strong inverse relation between tariff reduction and higher trade levels across Asia and the Pacific. Also, reverse causality between trade and inequality appears to work here, since more unequal nations influence trade positively. This regression does not take per capita income as a regressor since it tends to influence other explanatory variables strongly. However, infrastructure stock as well as quality does not appear to have any significant influence on trade. The country-specific effect does matter significantly with maximum effect being shown for Malaysia and minimum effect for India. Incidentally, Malaysia is one of the more open economies and it is not surprising that apart from tariffs, this country has adopted specific pro-trade policies to boost its trade.

Second, once the per capita income is introduced, it tends to capture most of the effects on trade. In columns 3–5 of Table 4, all regression models show that the richer the country, the higher the trade value. Tariffs play an important role and they always promote trade when reduced. However, the per capita income variable has a strong correlation with infrastructure variables, so it is not surprising that infrastructure stock appears to have retarded the trade of nations across the Asia-Pacific region. Under the circumstances, calculating the influence of infrastructure stock and quality in the absence of per capita income as a regressor, as is done in column 2,

Table 4 Trade, inequality and infrastructure: baseline regression results (dependent variable = ln trade GDP)

Variables	ln trade GDP	ln trade GDP	ln trade GDP	ln trade GDP
(1)	(2)	(3)	(4)	(5)
Income inequality (ln Gini)	0.443[b] (0.207)	−0.438[a] (0.262)	0.101 (0.261)	−0.106 (0.239)
Infrastructure stock (ln infrastock)	0.256 (0.196)	−0.462 (0.287)	−0.582[b] (0.260)	−0.422[a] (0.254)
Infrastructure quality (ln tdloss)	−0.861 (0.075)	−0.052 (0.081)	−0.069 (0.073)	−0.047 (0.071)
Per capita income (ln pc GDP)		0.722[c] (0.066)	0.391[c] (0.110)	0.403[c] (0.104)
Land (in land)			−0.974[c] (0.197)	
Trade liberalization (ln tariffs)	−0.312[c] (0.036)			−0.223[c] (0.041)
Observations	120	120	120	120
Adjusted R-squared	0.943	0.936	0.948	0.950
Country fixed-effects	Yes	Yes	Yes	Yes
Model	OLS	OLS	OLS	OLS
Breusch and Pagan LM test, chi2 (*p*-value)	82.41[c] (0.000)	102.80[c] (0.000)	104.44[c] (0.000)	85.70[c] (0.000)
Hausman test, chi2 (*p*-value)	12.22[b] (0.016)	11.87[b] (0.018)	32.23[c] (0.000)	10.32 [a] (0.067)
Fixed-effect coefficients	Max = 3.856[c] (Malaysia)	Max = 2.434[a] (Viet Nam)	Max = 5.276[c] (Malaysia)	Max = 3.732[c] (Viet Nam)
	Min = 2.526[b] (India)	Min = 0.101 (Rep. of Korea)	Min = 0.798 (Hong Kong, China)	Min = 2.312[b] (Rep. of Korea)

Note Robust standard errors in parentheses
[a]$p < 0.1$, [b]$p < 0.05$, [c]$p < 0.01$

appears reasonable. However, infrastructure does not appear to play any significant role in increasing openness of a country; of greater importance is the impact of tariffs, which boost trade with or without infrastructural development. This is an important finding.

Third, a significant positive influence of country-specific policies on trade openness apart from the influence of tariffs is again found. It is also not surprising that Malaysia and Viet Nam stand out as the most important countries with regard to having pursued individual policies vigorously. Clearly, tariffs, inequality and per capita income are found to be common variables significantly influencing trade, but it is also asserted that there are country-specific policies that matter for the trade openness. However, not only are the richer countries more open, the more economically unequal nations also influence trade across countries in the Asian and Pacific region at different levels of development. In fact, this is consistent with other studies that show trade variables such as exports are influenced by different factors, e.g. the real exchange rate and world demand (Marjit and Raychaudhuri 1997).

Table 5 Arellano–Bover dynamic panel data estimation: dependent variable equals inequality (ln Gini)

Variables	Coefficient	S.E.	*t*-statistic
Inequality (ln Gini)			
L1	0.120	0.122	0.978
L2	0.179[b]	0.089	2.005
Trade openness (ln trade GDP)	−0.022	0.056	−0.406
Infrastructure stock (ln infrastock)	−0.060	0.076	0.780
Infrastructure quality (ln tdloss)	−0.041	0.039	−1.035
Hansen Sargan J statistic, chi2 (*p*-value)	6.164 (0.723)		
Observations	78		
Instrument rank	14		

Note Dynamic panel counts White period instrument weighting matrix, White period standard errors and covariance (d.f. corrected). The estimation uses orthogonal deviation
[b]$p < 0.05$
L1 and L2 equal lags 1 and 2, respectively

Since the study highlights the role of infrastructure, real exchange rates and world income are dropped from the regressions to avoid multicollinearity problems. The impact of these dropped variables are captured in the fixed effect coefficients which are significant in most cases.

With regard to robustness checks, in order to rule out the possibility of endogeneity in Eqs. (2) and (3) above, a dynamic GMM estimator (system GMM) was used to analyse changes across countries and over time.[14] Also, the Arellano–Bover estimates presented in Tables 5 and 6 remove the weak instrumental variables and poor efficiency problems since they utilize more moment conditions. Table 5 provides Arellano–Bover estimates when the dependent variable is income inequality, whereas Table 6 presents the similar estimates when trade openness is the dependent variable. To test the appropriateness of the instruments used, the Sargan J statistics of over identifying restrictions in Tables 5 and 6 were used. The J statistics show that the applied instruments are valid. It should also be noted that the authors attempted to maintain some parity with the static and dynamic panel estimation, since that makes it easier to compare and contrast the results obtained.

It is apparent that past inequality significantly determines the present level of inequality (two periods of lagged inequality are statistically significant), but trade openness together with all the infrastructure variables have little influence on the movement of inequality in countries over time. In support of the previous findings (Table 3), system GMM estimates suggest persistence of inequality, since the initial level of inequality appears to be the only instrument that matters in the evolution of income inequality over space and time.

In contrast, one period of lagged values of trade openness significantly influences the current trade to GDP ratio, again showing persistence; however, unlike in

[14]Refer, for example Arellano and Bover (1995), Blundell and Bond (1998).

Table 6 Arellano–Bover dynamic panel data estimation: dependent variable equals trade openness (ln trade GDP)

Variable	Coefficient	S.E.	*t*-statistic
Trade openness (ln trade GDP)			
L1	0.373[b]	0.142	2.627
L2	0.101	0.100	1.015
Inequality (ln Gini)	0.333*	0.192	1.739
Trade liberalization (ln tariffs)	−0.157[a]	0.049	−3.180
Infrastructure stock (ln infrastock)	0.452	0.701	0.645
Infrastructure quality (ln tdloss)	−0.085	0.215	−0.393
Hansen Sargan J statistic, chi2 (*p*-value)	9.830 (0.198)		
Observations	65		
Instrument rank	13		

Note Dynamic panel counts White period instrument weighting matrix, and White period standard errors and covariance (d.f. corrected). The method uses orthogonal deviation
[a]$p < 0.01$, [b]$p < 0.05$
L1 and L2 equal lags 1 and 2, respectively

the earlier case for inequality, tariff liberalization together with inequality play a significant role in trade openness. This again supports the static panel result. Thus, although a more open economy does not significantly reduce inequality, the reverse is not true. However, infrastructure improvement is again not an important factor for the evolution of more trade and less inequality in the economy.

5 Conclusions

The foregoing analysis provides a synoptic view of the role of infrastructure in economic growth and income inequality as well as international trade of countries in Asia and the Pacific. The point that is emphasized throughout this paper is that infrastructure mainly appears as a complementary factor in the standard literature. As a result, the authors have discussed the role of infrastructure in a somewhat compartmentalized manner. In the existing literature on the subject, although not comprehensive or specific in terms of the theme under consideration in this paper the positive impact of infrastructure on growth is proved through growth regressions; however, the important point to note is that infrastructure (or public capital) is a factor whose efficiency is as equally important as its quantity. Possibly this applies to all factors of production, but for infrastructure this is emphasized time and again. The developing countries unambiguously show this positive impact, unlike some of the developed countries.

Unfortunately, most of this work does not address either trade or income distribution issues. The literature that deals with inclusive growth and poverty highlights the fact that accessibility to infrastructure, like rural roads or electricity, does not benefit the poor much. Hence, the result may be an unintended widening of income disparity. The authors suggest direct targeted policy intervention in this case.

The literature on infrastructure's role in growth or inclusive growth rarely discusses the trade issues at the same time. Given the $2 \times 2 \times 2$ structure of trade models, the role of infrastructure is discussed mainly as a trade-enabling or facilitating device as well as a facet that could reduce trade costs if used more efficiently. The existing literature on infrastructure and trade more or less concentrates on these two themes, without giving much attention to growth or inclusive growth in an integrated framework.

Going through the above literature, it is clear that one needs a study that looks at the individual causalities in an integrated framework and discusses the role of infrastructure in growth and income distribution in an open economy framework. This would look at infrastructure, in terms of both quantity and quality, as an important causal factor promoting trade and inclusive growth in the economy. The panel regressions carried out here fulfil this gap to some extent. However, it must be noted that country-specific studies are essential for gaining a better understanding of this issue, as is clearly revealed by the regression results.

In the panel regressions detailed in this paper, both infrastructure quantity and quality are found to affect inequality. However, the same is not true for trade. Infrastructure is overshadowed by trade liberalization measures when it comes to increasing trade in an economy. In addition, country-specific policies are shown to be significant determinants of both trade and inequality, apart from infrastructure development. Thus, for trade, it may be the exchange rate or foreign exchange management policies as well policies of protection to specific goods. For inequality, it may be some directly targeted poverty alleviation policies. Therefore, although infrastructure development may reduce inequality, this may be not so important for raising trade volumes. Also, persistence is important for both inequality and trade, pointing to the fact that initial levels matter and this cannot be overshadowed immediately by a sudden change in infrastructure or country-specific policy designs. This is the message that comes out from the unbalanced panel regressions, spanning 1975–2006, which were run for 14 countries at different levels of development in the Asia-Pacific region.

Infrastructure development unambiguously decreases inequality, but its effect on trade is not particularly significant. At the same time, there is no substitute for country-specific studies in understanding such linkages, since such country-specific policies together with infrastructure development are significant determinants in achieving higher trade accompanied by lower income inequality.

Acknowledgments An earlier version of the paper was presented at the international workshop 'Trade, Investment and Regional Integration: Lessons for Policy Makers', organized by ARTNeT/ESCAP and the Ministry of Commerce & Industry, Government of India, held at New Delhi, 11–12 March 2010. Authors sincerely acknowledge the useful comments of anonymous referee of ARTNeT/ESCAP, Mia Mikic and workshop participants. This paper was prepared as part of the ARTNeT initiative. The technical support of the United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP) is gratefully acknowledged. An earlier version of the paper was published as ARTNeT working paper no. 8210 in 2010. The opinion, figures and

estimates are responsibility of the authors and should not be considered as reflecting the views or carrying the approval of the United Nations, ARTNeT, RIS and JU. Any remaining errors are the responsibility of the authors.

Appendix: List of Variables and Sources

Label	Content	Sources
Inequality (INQ)	Gini coefficient	UNU-WIDER World Income Inequality Database, Version 2.0c, May 2008 (UNU-WIDER, 2008, and ADB, 2007)
Trade openness (TO)	Trade (export and import) as percentage of GDP (per cent)	World Development Indicators CD-ROM 2009, (World Bank 2009)
Trade liberalization (TL)	Simple average tariff (per cent)	World Trade Integrated Solution (WITS)
Infrastructure stock (IS)	Infrastructure quantity, which is summed over normalized indicators such as (a) air freight transport (million tonnes/km); (b) electric power consumption (kWh per capita); (c) fixed line and mobile phone subscribers (per 1,000 people); (d) households with television (percentage of total households); (e) railway lines (km per 1,000 km^2 of geographical area); (f) roads, (km per 1,000 km^2 of geographical area); and (g) personal computers (per 1,000 people)	World Development Indicators CD-ROM 2009, World Bank (2009)
Infrastructure quality (IQ)	Electric power transmission and distribution losses (percentage of output)	World Development Indicators 2009 online database (World Bank 2009)
Per capita income (PCI)	GDP per capita (constant 2000 US$)	
Geography (Land)	Per capita land availability (km^2)	
Inflation (Inf)	GDP deflator (annual percentage)	
Population (Pop)	Working population (population ages 15–64, percentage of total population)	

References

Ali I, Pernia EM (2003) Infrastructure and poverty reduction—what is the connection? ERD Policy Brief No. 13. Asian Development Bank, Manila

Anderson JE, van Wincoop E (2004) Trade costs. J Econ Lit XLII(3):691–751

Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev Econ Stud 58(2):277–297

Arellano M, Bover O (1995) Another look at the instrumental variable estimation of error component models. J Econ 68(1):29–51
Asian Development Bank (2007) Key indicators of developing Asian and Pacific Countries. Manila
Barro RJ (1990) Government spending in a simple model of endogenous growth. J Polit Econ 98 (5):S102–S125
Barro RJ (1991) Economic growth in a cross-section of countries. Quart J Econ 106:407–443
Bhagwati JN, Srinivasan TN (2003) Trade and poverty in the poor countries. www.econ.yale.edu/~srinivas/trade_poverty.pdf
Blundell R, Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. J Econ 87(1):115–143
Brenneman A, Kerf M (2002) Infrastructure and poverty linkages: a literature review. World Bank, Washington, D.C
Briceño-Garmendia C, Estache A, Shafik N (2004) Infrastructure services in developing countries: access, quality, costs and policy reform. Working Paper No. 3468. World Bank, Washington D.C
Calderón C, Servén L (2005) The effects of infrastructure development on growth and income distribution. Policy Research Working Paper 3643. World Bank, Washington, D.C
Campano F, Salvatore D (2007) Economic development and income distribution. J Policy Model 29:553–566
Canning D (2006) A database of world infrastructure stocks 1950-95. Policy Research Working Paper No. 1929. World Bank, Washington D.C
Datt G, Ravallion M (2002) Is India's economic growth leaving the poor behind? Policy Research Working Paper No. 2846. World Bank, Washington, D.C
Deardorff AV (2001) International provision of trade services, trade, and fragmentation. Rev Int Econ 9(2):233–248
Edwards S (1997) Trade policy, growth, and income distribution. Am Econ Rev 87(2):205–210
Economic and Social Commission of Asia the Pacific (ESCAP) (2001) Reducing disparities: balanced development of urban and rural areas and regions within the countries of Asia and the Pacific. Bangkok
Economic and Social Commission of Asia the Pacific (ESCAP) (2009) Asia-Pacific trade and investment report 2009. United Nations, Bangkok
Fan S, Zhang L, Zhan X (2002) Growth, inequality and poverty in China: the role of public investments, research report 123. International Food Policy Research Institute, Washington, D.C
Gourdon J, Maystre N, de Melo J (2008) Openness, inequality and poverty: endowments matter. J Int Trade Econ Dev 17(3):343–378
Hulten CR (1996) Infrastructure capital and economic growth: how well you use it may be more important than how much you have. NBER Working Paper No. 5847. National Bureau of Economic Research, New York
Levine R, Renelt D (1991) A sensitivity analysis of cross-country growth regressions. Policy Research Working Paper 609. World Bank, Washington, D.C
Marjit S, Raychaudhuri A (1997) India's exports. Oxford University Press, New Delhi
Meschi E, Vivarelli M (2009) Trade and income inequality in developing countries. World Dev 37 (2):287–302
Pinstrup-Andersen P, Shimokawa S (2006) Rural infrastructure and agricultural development. Paper presented at the World Bank ABCDE Conference, Tokyo
Roland-Holst D (2006) Infrastructure as a catalyst for regional integration, growth, and economic convergence: scenario analysis for Asia. ERD Working Paper 91. Asian Development Bank, Manila

Slaughter MJ (1997) Per capita income convergence and the role of international trade. Am Econ Rev 87(2):194–199

Straub S (2008) Infrastructure and development: a critical appraisal of the macro level literature. Policy Research Working Paper 4590, World Bank, Washington, D.C

Winters LA, McCulloch N, McKay A (2004) Trade liberalization and poverty: the evidence so far. J Econ Lit XLII:72–115

World Bank (1994) World development report 1994. Washington, D.C

World Bank (2009) World development indicators CD-ROM. Washtington, D.C

A Theoretical Model of Trade, Quality of Health Services and Signalling

Kausik Gupta and Tonmoy Chatterjee

Abstract Export of quality of health services to the foreign consumers is considered as a popular way of health trade in the South in recent years. Trade in quality of health services may suffer from the problem of quality signalling from the point of view of the patients of the North. Actually, patients of the North are misdirected by the quality of health services if proper signalling is not provided by the health quality producers of the South. Again signalling incurs some cost and hence production and export of high quality health services becomes unprofitable to the health quality producers of South. Hence, there exists a definite demand–supply mismatch regarding health quality. To capture the issue of quality signalling in the context of health services we want to develop a partial equilibrium framework, where the producer of the health quality in South, we refer to it as Multinational Health Service Provider (MNHSP), may enjoy some amount of monopoly power as this MNHSP has some special skills or it may introduce new quality of health services. The model shows that a movement from autarky to no autarky situation leads to, under reasonable conditions, a decline in signalling costs associated with high quality health services and also leads to an overall increase in demand for high quality health services in the South.

This paper is a part of the doctoral dissertation of Tonmoy Chatterjee, which is in progress for PhD degree of Rabindra Bharati University, Kolkata, India. The authors are indebted to an anonymous referee, Prof. Sugata Marjit and Prof. Tarun Kabiraj for their valuable comments on an earlier draft of this paper. This paper has also benefitted from the comments of seminar participants at Rabindra Bharati University (2015). However, errors that may remain are the sole responsibility of the authors.

K. Gupta (✉)
Department of Economics, Rabindra Bharati University, Kolkata, India
e-mail: kausik2k1@rediffmail.com

T. Chatterjee
Department of Economics, Ananda Chandra College, Jalpaiguri, India

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_13

Keywords Health quality · International trade · Dynamic game · Signalling cost

JEL Classification I11 · F12 · F23 · C73 · D82

1 Introduction

Healthcare has emerged as one of the most progressive and largest service sectors in India. India's medical tourism sector is expected to experience an annual growth rate of 30 %, making it a Rs. 9,500 crore industry by the end of 2016. Indian Brand Equity Foundation (IBEF) suggests that the Indian healthcare sector is expected to become a US $280 billion industry by 2020 with spending on health estimated to grow 14 % annually.

Multilateral liberalization of trade in goods has taken place for the past half century. However, liberalization of trade in services began only in 1994 with the end of Uruguay Round. The General Agreements on Trade in Services (GATS) is the first and the only set of multilateral rules covering international trade in services. The agreement was concluded at the Uruguay Round (1986–1993). It came into force in January 1995. The GATS has argued to gradually liberalize and expand trade in the service sector. It has a *"built in agenda"* mandating members to kick off progressive liberalization negotiation on services from January 2000. The Guidelines and Procedures of Negotiations were adopted by the Council for Trade in Services on March 2001, as provided in the GATS Article XIX[1]: 3. It seems useful to mention here that through the GATS, the member states of the World Trade Organization (WTO) have the option to commit themselves to liberalize health services. Article I of the General Agreement on Trade in Services (GATS), following Bhagwati (1984) and Sampson and Snape (1985), discusses a four part typology of trade in health services.[2] It has been rightly pointed out in GATS that trade in health services may occur through four modes and they are: (1) *cross-border supply:* where the service is provided remotely from one country to another, such as telemedicine via Internet or satellite, or international health insurance policies; (2) *consumption abroad:* where individuals use a service in another country, such as patients travelling to take advantage of foreign health care facilities; (3) *commercial presence:* where a foreign company sets up operations within another country in order to deliver the service, such as hospitals, health clinics or insurance offices and (4) *presence of natural persons:* where individuals such as doctors, nurses or midwives travel to another country to supply a service there on a

[1]See Kelegama (2009) for details.

[2]Interested readers can go through the papers by Bhagwati (1984) and Sampson and Snape (1985). See also Findlay and Warren (2000).

temporary basis. The majority of the patients coming to India seek for good cardiac and cancer treatments. Orthopaedic and kidney-related treatment also forms a significant percentage. A little bi-modality among cardiology patients, focusing under 19 and middle-aged patients, has been noticed in the work of Banik et al. (2010). Mudu (2004) has strongly argued for export in health services along with tourism. In his article, he has shown that whether it is for cataract or cardiac surgery, dental treatment or hip replacement operations, treatment cost in India is a fraction of what an identical procedure would cost in Europe or North America. Again large hospitals in Mumbai, Chennai and New Delhi have long been receiving patients from neighbouring South Asian and Gulf countries. They are now trying to attract patients from Africa, Europe and North America, marketing themselves as centres capable of delivering world class medical services at low cost. Similarly from the paper by Turner (2007) we can say that India, Indonesia, Malaysia, the Philippines, Singapore, Thailand and many other countries market themselves as major destinations for 'medical tourism'. Health-related travel, once promoted by individual medical facilities such as Bumrungrad International Hospital and Bangkok International Hospital, is now driven by government agencies, public–private partnerships, private hospital associations, airlines, hotel chains, investors and private equity funds and medical brokerages. 'Medical tourists' include patients who usually try to avoid treatment delays and opt for timely access to health care. Medical travellers also include uninsured Americans and other individuals unable to afford health care in their home settings. Destination nations regard medical tourism as a resource for economic development. Affordability combined with inexpensive air travel, low cost telecommunications, digitized patient records, widespread access to information through the internet and an increasingly sophisticated medical travel industry to manage all these processes on the patient's behalf causes medical tourism as an alternative for uninsured or underinsured individuals (Turner 2007).

Again Mudu (2004), Banik et al. (2010), Turner (2007) and Menck (2005) have rightly pointed out the proper reasons due to which patients of North avoid to go to the South to get cheaper treatment. Overall they have claimed that the patients of North used to suffer from the phobia that they may be misdirected in the South regarding health quality. It is to be noted that improving patient safety, medication, infection prevention and control, quality performance and improvement in the environment of health care is a primary concern for hospitals and medical facilities involved in medical tourism and these hospitals and health care providers voluntarily seek accreditation of Joint Commission International (JCI) to attract patients from all over the world. In India Quality Council of India (QCI), an organization of Government of India, has set up National Accreditation Board for Hospitals and Healthcare Providers (NABH). In a NABH accredited hospital, there is strong focus on patient rights and benefits, patient safety, control and prevention of infections in hospitals and practicing good patient care protocols like special care for vulnerable groups, critically ill patients and better and controlled clinical outcome. There are 16 JCI accredited and 63 NABH accredited healthcare providers in India. The Indian government is taking steps to address infrastructure issues that hinder the country's growth in medical tourism. Most estimates claim treatment costs in India start at around a tenth of the price of comparable treatment in America or Britain.

It is quite clear from above that foreign patients prefer Indian health care as their destination for better treatment. In spite of the presence of few obstacles in front of Indian health service providers, they have tried their best to cater the markets of foreign patients but still the proportion of export of health services as percentage of global foreign consumer of health services is quite low. Poor infrastructure of domestic health services, lack of purchasing power for better health services of domestic patients and obviously last but most important among others the information asymmetry regarding the quality of health services between domestic health service provider and patients from North are some of the obstacles, which are prevailing in most of the developing economies. In this paper, we want to locate and try to solve the problem of information asymmetry associated with health care quality (following Spence (1974) and Wolinsky (1983))[3] between health service provider of South and consumer of health quality from North. To solve the problem of information gap health service provider of South can introduce the mechanism of quality signalling in health services. In order to capture the issue of quality signalling in the context of health services initially we have developed a partial equilibrium framework at autarky, where the producer of the health quality in South, we refer to it as Multinational Health Service Provider (MNHSP), may enjoy some amount of monopoly power as this MNHSP has some special skills or it may introduce new quality of health services. Then we form a strategic game and solve it for the preferences and prices of high and low quality of health services through backward induction. Finally, we break the autarky and move towards the regime of international trade through the inflow of foreign patients from the North to the South and have tried to examine the role of health quality signalling in this present regime.

The issue of vertical differentiation model with a monopoly has been studied recently by Acharyya (2000). Using a linear utility function (as in Tirole 1988), he has examined how a change in the income distribution can affect the quality choice of a monopolist. He has proved that income redistribution cannot only increase but also decrease the optimal quality provided by the monopolist. However, unlike in most of the vertical differentiation models, Alonso and O'Donnell (2001) have not been concerned with analyzing the strategic interaction between firms in the presence of income disparities. They have also analyzed the strategic interactions between a unique innovative firm which faces a competitive fringe in the established quality and a health care system which attempts to use income redistribution as a tool for increasing access to the health care innovation. In our paper, we use a log linear characterization of the utility of consumers which is closer to the work of

[3]Following Spence (1974) and Wolinsky (1983), in this paper the representative MNHSP of South can signal its quality of health services to the patients using high price and quantity restrictions, such that patients' information gap can be solved.

Alonso and O'Donnell (2001). Apart from that, we have used a simple income distribution function which allows us to correlate individual's utility with prices, quality and the consumer's income.

Though there exists quite a few theoretical works based on issues like monopoly health care, quality innovation and international trade, but unfortunately none of these papers have considered the issue of asymmetric information in the context of trade through health services. The present paper has tried to fill up the lacuna from the existing literature. Moreover, in this paper we have shown that international trade in health services can reduce the cost of signalling of signalling of health care quality and in turn it can drive out lower quality of health care services from the market.

The paper is organized in the following manner. Section 2 discusses the rationale behind signalling cost. Section 3 discusses the model under autarky. The model with trade has been discussed in Sect. 4. Finally, the concluding remarks are made in Sect. 5.

2 Rationale Behind Signalling Cost

To fix ideas we start from a very brief rationale behind the concept of signalling cost. We start from a situation when there is no trade and we focus only on a developing economy which we shall refer to as 'South'. Suppose in the 'South' there exists a Multinational Health Service Provider (hereafter MNHSP) which has *two* broad faces[4]: one is MNHSP which provides high quality health care services and the other is the MNHSP which pretend to provide high quality health care services but actually provide low quality health care services. The MNHSP in the 'South' enjoys some amount of monopoly power, as it has some special skills or it may introduce new quality of health services, and can thus provide profit maximizing level of quality of a monopoly health service provider. We consider high quality health care as Q_h^H (we shall refer to it as 'H' quality) and low quality health care as Q_h^L (we shall refer to it as 'L' quality). The respective prices of the above mentioned two types of quality are denoted as P_H and P_L. Here, we assume quantity demanded for H quality health services is q^H and quantity demanded for L quality health services is q^L and we also assume that $q^H < q^L$, as price of high quality health services P_H is greater than the price of low quality health services P_L.

[4]By two broad faces of the MNHSP we mean that though the representative MNHSP is basically a high quality health service provider of South but in reality this health service provider can produce both kinds of health care quality and hence the patients become unaware of the fact that which quality of health services will be provided by the MNHSP. It is to be noted that the MNHSP can signal its high quality health services by charging high price and restricting supply of its high quality health services to the patients. Those who will be unable to afford the high quality health services will consume low quality health care services either from a MNHSP which pretends to be a low quality health service provider or from other low quality health service providers of South.

Suppose we denote by $C_j(Q_h{}^j)$ (for $j = L, H$) as the unit cost for providing jth quality of health care service. In the South, it happens frequently that a low quality MNHSP sells its health care services at high quality prices, P_H, pretending its services as a high quality one. In this case the corresponding quantity demanded is q^H. Thus at (P_H, q^H), the MNHSP's profit will be

$$\left[P_H - C_L\left(Q_h^L\right)\right]q^H = \pi^L(P_H, q^H) \tag{1}$$

If the MNHSP consider full market coverage then it will have to sell at price P_L and the MNHSP's profit will be

$$\left[P_L - C_L\left(Q_h^L\right)\right]N = \pi^L(P_L, N) \tag{2}$$

Here, N is the total number of patients in the South and it is exogenously given. It is to be noted that when the low quality producer is indifferent between (P_H, q^H) and (P_L, N) from the point of view of profit we have

$$\left[P_H - C_L\left(Q_h^L\right)\right]q^H = \left[P_L - C_L\left(Q_h^L\right)\right]N$$

from which we get

$$q^H = \left[P_L - C_L\left(Q_h^L\right)\right]N/\left[P_H - C_L\left(Q_h^L\right)\right] \tag{3}$$

Equation (3) implies low quality MNHSP can choose P_H and q^H as the profit maximizing price and quality respectively. It also implies that as N increases q^H increases.

Consumers in the South frequently select wrong quality health services due to asymmetric or imperfect information. To overcome this situation MNHSP has to signal its quality. To signal its own quality the MNHSP has to incur a cost, we can refer it as *signalling cost*. This signalling cost occurs due to the information gap about provision of high quality services by MNHSP.

If a high quality MNHSP produces at (P_H, q^H) under imperfect information, its profit will be

$$\left[P_H - C_H\left(Q_h^H\right)\right]q^H = \pi^H(P_H, q^H) \tag{4}$$

Profit under perfect information is given by

$$\left[P_H - C_H\left(Q_h^H\right)\right]q^{H*} = \pi^H(P_H, q^{H*}) \tag{5}$$

where q^{H*} is the maximum possible quantity demanded for high quality health services and we assume $q^{H*} > q^H$.

Thus the cost of signalling is obtained by considering the fact that profit under perfect information is greater than profit under imperfect information regarding high quality health services, i.e. $[P_H - C_H(Q_h^H)]\ q^{H*} > [P_H - C_H(Q_h^H)]q^H$, which actually implies

$$\left[P_H - C_H\left(Q_h^H\right)\right]\left(q^{H*} - q^H\right) > 0 \tag{5.1}$$

It implies signalling cost is positive[5] and decreases as N increases. Intuitively we can show that an increase in the number of patients due to the movement of foreign patients from North to south to get high quality treatment may enhance the demand for high quality health services of South-based MNHSP. A rise in quantity demanded of high quality health treatment leads to an increase in the profit level of our representative MNHSP in the absence of signalling. This rise in profit of high quality health services may induce the heath service provider to provide more of high quality health treatment and hence it leads to minimize the gap between profit under perfect information and profit under imperfect information regarding high quality health services. We thus have the following proposition.

Proposition 1 *The MNHSP faces a positive signalling cost and moreover, this cost falls as N increases.*

3 The Model Under Autarky

In this section, we emphasize on the determination of price and quantity of our health care quality producing MNHSP under autarky, i.e. we consider here only the South. To analyze the model we have considered a dynamic game. The timing of the game is as follows. Here we consider a two-stage game. In the first stage, the health care decides price of innovation regarding health services and also takes care about international market coverage. In the second stage, the consumers/patients in both countries select/choose the quantity of quality for health services. The solution is obtained by backward induction.

[5]One can denote the value of q^H in Eq. (3) as $\bar{q}^H$ and $\bar{q}^H$ is shown explicitly as a function of N. Again one can denote the optimal value of q^H in Eq. (4) as $\tilde{q}^H$ and $\tilde{q}^H$ can be expressed as a function of N. Though the exact $\tilde{q}^H$ cannot be shown explicitly but it is obviously a positive function of N, i.e. $\tilde{q}^H = \tilde{q}^H(N)$. We can consider $q^H(N) = \bar{q}^H(N) + \tilde{q}^H(N)$. Given the fact $q^{H*} > q^H(N)$ we can rewrite the signalling cost in Eq. (5.1) as $[P_H - C_H(Q_h^H)]$ $\{q^{H*} - q^H(N)\} > 0$. It implies Proposition 1 is valid. It is also to be noted that as quality is given $C_H(Q_h^H) = C_H$ (fixed). In fact $C_L(Q_h^L) = C_L$ is also fixed. So, our results will remain unchanged if we interpret signalling cost by comparing value of q^H under perfect information with the value of q^H under imperfect information, i.e. $P_H q^{H*} > P_H q^H(N)$, or, $P_H[q^{H*} - q^H(N)] > 0$. In this case also the Proposition 1 is valid.

3.1 Stage 2: Patients Select Quality of Innovative Health Services

Here we assume income of a representative patient in the South as Y and it is assumed that Y is uniformly distributed between Y^M and 0. Then probability density function of income distribution in the South can be written as

$$f(Y) = \begin{cases} N_{\mathrm{R1}} & \text{for } Y^M \geq Y \geq Y^H \\ N_{\mathrm{B1}} & \text{for } Y^H > Y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

where N_{R1} and N_{B1} are, respectively, the frequencies of rich and poor in the South, Y^H is the critical level of income at which patients will be indifferent between high and low quality, that is, at least Y^H amount of income is required to afford high quality health services, again income required to buy high quality health care services and Y^M is highest level of income of the society. Thus $N_{\mathrm{R1}}(Y^M - Y^H)$ and $N_{\mathrm{B1}}Y^H$ are the income of rich and poor, respectively, in the South.

The representative consumer's utility function can be written as[6]

$$V(Y - e(Q_h), Q_h) = \ln U[Y - e(Q_h), Q_h] = \ln[Y - e(Q_h)] + \ln Q_h \tag{6}$$

Here, $V(Y - e(Q_h), Q_h)$ implies utility derived from both quality of health services and also as a result of expenditure on other commodities. The expenditure on health quality Q_h is given by $e(Q_h)$. The first component of the utility function on the right hand side of Eq. (6) implies utility derived from commodities other than quality of health services. Second component of the utility function on the right hand side of Eq. (6) implies utility derived from the quality of health services. To obtain the proper level of health quality, the patients take into account of their expenditure for different levels of health care quality. It is to be noted that degree of health care quality is positively associated with its prices and hence we can have the following specifications:

Expenditure for no treatment: $e_0(Q_h^0) = 0$.
Expenditure on treatment with quality less than $Q_h^H = e_L(Q_h^L) = P_L$
Expenditure on treatment with quality $Q_h^H = e_H(Q_h^H) = P_H$.

[6]This specification follows from the expected utility function as used by Goddeeris (1984a, b). In the context of health services it can be specified as $EU = \alpha U^1(Y) + (1 - \alpha)\, U^2(Y - e(Q_h), Q_h)$, where α and $(1 - \alpha)$ are, respectively, the probability of occurrence of U^1 and U^2 utility functions. This type of utility function is widely used in the literature on quality of health services. Consideration of such type of utility function implies the following: (i) willingness to pay for health care quality increases with income and (ii) preferences will be of convex. For details see Alanso and O'Donnell (2001). For our analysis, we have considered only the $U^2(Y - e(Q_h), Q_h)$ utility function and has denoted it by $V(Y - e(Q_h), Q_h)$.

It is to be noted that Q_h^H is the threshold level of health quality and $0 < P_L < P_H$. It is also to be noted that $Q_h^H > Q_h^L$ as the level of high quality health care is always greater than the level of low quality health care.

We can rewrite the utility function $V(Y - e(Q_h), Q_h)$ *for all patients* (instead of only the representative consumer) for two categories. The first category is the utility function of all patients who spend on high quality health services. The second category is the utility function of all patients who spend on low quality health services. We assume that for both the categories the patients spend on health quality services and also on other commodities. The specifications of the utility functions for a patient are as follows:

$$V\left(Y - e_H\left(Q_h^H\right), Q_h^H\right) = \ln U\left[Y - e_H\left(Q_h^H\right), Q_h^H\right] = \ln\left[Y - e_H\left(Q_h^H\right)\right] + \ln Q_h^H \quad (7)$$

$$V\left(Y - e_L\left(Q_h^L\right), Q_h^L\right) = \ln U\left[Y - e_L\left(Q_h^L\right), Q_h^L\right] = \ln\left[Y - e_L\left(Q_h^L\right)\right] + \ln Q_h^L \quad (8)$$

Equation (7) corresponds to the utility function of a patient for high quality health services and Eq. (8) corresponds to the utility function of a patient for low quality health services.

Let Y^H be the threshold level of income where patients are indifferent between high and low quality of health services. In this case, we compare Eqs. (7) and (8) by equating them and also using the fact $e_L(Q_h^L) = P_L$ and $e_H(Q_h^H) = P_H$ we get

$$Y^H = \left(Q_h^H P_H - Q_h^L P_L\right) / \left(Q_h^H - Q_h^L\right) \quad (9)$$

Here $Y^H > 0$ as $Q_h^H > Q_h^L$ and $P_H > P_L$.

It can be shown that for all $Y \in (0, Y^H)$ the patients will prefer low quality health services, whereas for all $Y \in (Y^H, Y^M]$ the patients will prefer high quality health services. This is explained with the help of the Fig. 1.

Figure 1 has been drawn on the basis of Eqs. (7) and (8). Using the fact that $e_H(Q_h^H) = P_H$ one can rewrite Eq. (7) as

$$U\left(Y - e_H\left(Q_h^H\right), Q_h^H\right) = -P_H Q_h^H + Y\, Q_h^H = U\left(Y, Q_h^H\right) \quad (7.1)$$

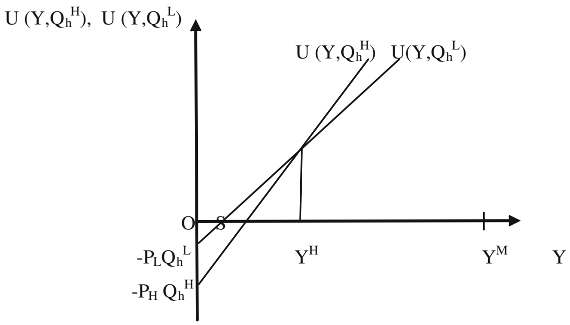


Fig. 1 Determination of threshold level of income

Similarly using the fact that $e_L(Q_h^L) = P_L$ one can rewrite Eq. (8) as

$$U\left(Y - e_L\left(Q_h^L\right), Q_h^L\right) = -P_L Q_h^L + Y\, Q_h^L = U\left(Y, Q_h^L\right) \tag{8.1}$$

From Eq. (7.1), we find that the slope of $U(Y, Q_h^H)$ function is given by

$$(\mathrm{d}U/\mathrm{d}Y)_{\text{for high quality}} = Q_h^H$$

Again from Eq. (8.1) we find that the slope of $U(Y, Q_h^L)$ function is given by

$$(\mathrm{d}U/\mathrm{d}Y)_{\text{for low quality}} = Q_h^L$$

As $Q_h^H > Q_h^L$ the $U(Y, Q_h^H)$ function is steeper than the $U(Y, Q_h^L)$ function. In Fig. 1, patients whose income lies within the area OS will not buy any health service (the number of consumers who will prefer no treatment). It is to be noted that in the figure SY^H is the region, where $N_{B1}Y^H$ number of patients will consume low quality health services, whereas, Y^HY^M is the region where $N_{R1}(Y^M - Y^H)$ number of patients will consume high quality health services. Suppose due to adaptation of new technology or innovation in the field of medical sciences there is an enhancement of the existing level of high quality health services, which is a rise in the level of Q_h^H. This rise in Q_h^H causes a leftward shift of the $U(Y, Q_h^H)$ locus and hence we get an increase in the number of consumers of high quality health services as critical level of income to buy high quality health services become lower due to more availability of high quality of health services and a decline in the number of consumers of low quality health services as more consumers can afford high quality health facilities. It is to be noted that number of people with no treatment may also decline due to such type of exogenous shock on health quality.

Remarks 1 Given the prices of different level of health quality, an improvement in the existing high quality health services may enhance the demand for high quality health services and reduces the number of consumers with zero demand for health care quality.

We next pass on to stage 1 of the game.

3.2 Stage 1: Price Selection of the Profit Maximizing MNHSP

From stage 2 of the game, we have found that Y^H is the threshold level of income at which patients are indifferent between high and low quality health services. From our model we also find a range for the number of patients who will prefer low quality health services. However, we can say that $N_{B1}Y^H$ is the *maximum* number of patients who will prefer low quality health services.[7] The number of patients who will prefer high quality health services can be definitely determined and is given by $N_{R1}(Y^M - Y^H)$. Given the fact that patients usually prefer high quality health services, we assume that the number of patients who will prefer low quality health services is $N_{B1}Y^H$, because after Y^H level of income patients will switch over to high quality health services and until Y^H is achieved patients will opt for low quality health services. Throughout we have assumed a given level of quality of health care services and they are indexed in such a manner so that the following inequality $1 < (Q_h^H / Q_h^L) < 2$ is satisfied.

The degree of health care quality is positively associated with its cost of production and hence we can write the following chart:

Cost of no treatment: $C_O(Q_h^0)$.
Cost of quality less than $H = C_L(Q_h^L) = C_L$
Cost of quality $H = C_H(Q_h^H) = C_H$

It is to be noted that H is the threshold level of health quality and we assume that C_L and C_H are fixed as health care quality is given. We also assume that $0 < C_L < C_H$.

If $N_{B1}Y^H$ number of patient select low quality health services, for given quality, the MNHSP will produce low quality health services and will attempt to maximize its profit, i.e.

$$\begin{aligned}\underset{P_L}{\text{Max}}\, \pi_L(Q_h^L, Q_h^H, Y, P_H, P_L) &= P_L N_{B1} Y^H - C_L \\ &= P_L N_{B1}\left(Q_h^H P_H - Q_h^L P_L\right)/\left(Q_h^H - Q_h^L\right) - C_L\end{aligned}$$

If we set $\partial \pi_L / \partial P_L = 0$ we get

$$P_L = Q_h^H P_H / 2Q_h^L \tag{10}$$

Under the assumption of quality indexation of the form $1 < (Q_h^H / Q_h^L) < 2$ we find that Eq. (10) implies $P_L < P_H$.

If $N_{R1}(Y^M - Y^H)$ patients select high quality health services, for given quality, the MNHSP will produce high quality health services and will attempt to maximize its profit, i.e.

$$\begin{aligned}\underset{P_H}{\text{Max}}\, \pi_H(Q_h^L, Q_h^H, Y, P_H, P_L) &= P_H N_{R1}(Y^M - Y^H) - C_H \\ &= P_H N_{R1}\left[\, Y^M - \left\{\left(Q_h^H P_H - Q_h^L P_L\right)/\left(Q_h^H - Q_h^L\right)\right\}\right] - C_H\end{aligned}$$

If we set $\partial \pi_H / \partial P_H = 0$, we get using Eq. (10)

[7] We can also interpret Y^H as the minimum level of income at which patients will prefer high quality health services.

$$P_H = 2Y^M\left(Q_h^H - Q_h^L\right)/3Q_h^H \tag{11}$$

Using Eq. (11) in (10) we get

$$P_L = Y^M\left(Q_h^H - Q_h^L\right)/3Q_h^L \tag{12}$$

Equations (11) and (12) imply that the game is complete as the prices are selected by MNHSP. It is to be noted that using the fact $1 < (Q_h^H/Q_h^L) < 2$ one can check by comparing Eqs. (11) and (12) that $0 < P_L < P_H$.

4 The Model with Trade

We now consider trade in health services. For this purpose we consider the foreign country as a developed one and we refer to it as the 'North'. For North we assume that patients' income is Y^f and it is assumed to be uniformly distributed between Y^M and $2Y^H$. Then probability density function of income distribution in North can be written as

$$f(Y^f) = \begin{cases} N_{\mathrm{R2}} & \text{for } Y^M \geq Y^f \geq 2Y^H \\ 0 & \text{otherwise.} \end{cases}$$

where N_{R2} is the frequency of rich in the North, $2Y^H$ is income required to buy high quality health care services and Y^M is highest level of income of the society. Thus $N_{\mathrm{R2}}(Y^M - 2Y^H)$ is the income of rich in the North.[8]

Patients of the North will move to the South for their treatment if the following inequality holds

$$P_H^N \geq P_H \tag{13}$$

where P_H^N represents the price of high quality health services in the North. It is to be noted that patients of North can enjoy $U(Y^f - e_H^N(Q_h^H), Q_h^H) = -P_H^N Q_h^H + Y^f Q_h^H$ level of utility from the consumption high quality health services at North. Whereas, the same consumers can enjoy a higher level of utility, that is, $U(Y^f - e_H(Q_h^H), Q_h^H) = -P_H Q_h^H + Y^f Q_h^H$ from the consumption of same level of health quality at South.

It implies for given pdf of income distribution of north we get

[8]As income required to buy high quality health services in the South (Y^H) must be less than the income required to buy the same services in the North we have assumed that income for the North as 2 Y^H.

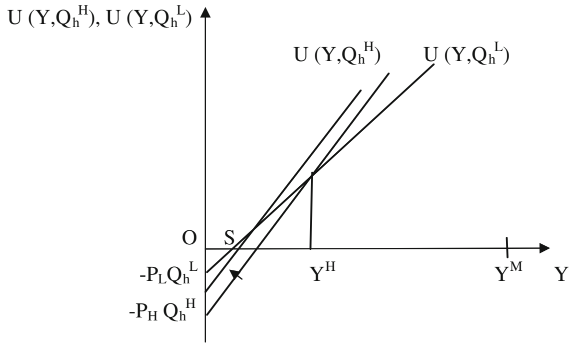


Fig. 2 Threshold level of income with declining signalling cost

$$
\begin{aligned}
U\left(Y^f - e_H^N\left(Q_h^H\right), Q_h^H\right) &> U\left(Y^f - e_H\left(Q_h^H\right), Q_h^H\right) \\
-P_H Q_h^H + Y\, Q_h^H &> -P_H^N Q_h^H + Y^f Q_h^H
\end{aligned}
$$

This expression basically justifies the rationality of the inequality (13). Thus, we can set the following inequality regarding prices of different levels of health care quality

$$0 < P_L < P_H \leq P_H^N \tag{14}$$

From (14) we can say that P_L is less than the price of high quality health services, that is, P_H. Here, P_H is greater than P_L basically due to the fact that to produce better quality health services MNHSP has to incur higher production cost, along with signalling cost, to capture the market of North along with the home market. As the value of P_H^N is higher compared to P_H, it implies that foreign patients of high quality health services will move from the North to the South for their treatment (as already mentioned in the context of inequality (13)) and hence the signalling cost to ensure the high quality health services will go down.[9] It is to be noted that a decline in the signalling cost due to increase in demand for high quality health services in the South will create a downward pressure on its price.[10] Moreover, a fall in P_H will shift the $U(Y, Q_h^H)$ schedule to the leftward direction, which in turn causes an increase in domestic demand for high quality health services. This is shown in Fig. 2.

The intuition behind the result can be explained on the basis of the fact that a move from a regime of autarky to no autarky may lead to an increase in probable demand for high quality health services, since the willingness to pay for high

[9]The reason has been already explained in Proposition 1.

[10]This result is quite compatible with the findings of Bagwell and Riordan (1991). Moreover, unlike Bagwell and Riordan (1991) in this paper we have interpreted trade in quality of health services from North to South as the main source behind the fall in the price level of high quality health services at South.

quality health services in the North is much higher relative to its Southern counterpart. It is to be noted that $P_H^N \geq P_H$ implies the patients of North enjoy higher level of utility from the consumption of high quality health services of South. As a consequence of which patients of North will move from North to South to cater for Q_h^H and hence the demand for high quality health services will go up. It is to be noted that as the number of consumers of quality Q_h^H increases the signalling cost associated with this quality will go down and therefore, P_H will decline further.[11] Fall in P_{H} again raises the domestic demand for high quality health services of the representative MNHSP. We summarize our results in the form of the following proposition.

Proposition 2 *A movement from a regime of autarky to no autarky may lead to*

(i) *a decline in the signalling cost associated with high quality health services of the South,*
(ii) *overall increase in the demand for high quality health services in the South. It increases not only for the movement of foreign patients (patients from the North) who demand such high quality health services but also due to an increase in the number of domestic patients (patients of the South) who demand the same high quality health services.*

5 Concluding Remarks

Health service sector and health intermediate sector (medical devices and equipments sector) are gaining importance among the economists and hence these sectors in recent years are considered as one of the most important parts of the social sector of any developing economy like India. In spite of such kind of importance among policymakers, this sector has not performed well especially in the context of delivery of health quality from South to North. There are many reasons which prevail to explain the issue of lack of trade in health services between the North and the South and one of them is the information asymmetry regarding the quality of health services among domestic health service provider and patients from North. That is, trade in quality of health services may suffer from the problem of quality signalling from the point of view of the domestic health service producers to the patients of the North. In fact the patients of the North are misdirected in this regard. To capture the above mentioned problem in the present paper we have considered a

[11]With international trade and in the absence of asymmetric information the profit function of our high quality MNHSP can be written as $\pi_H^{\mathrm{Trade}} = P_H^{\mathrm{Trade}}[N_{\mathrm{R1}}(Y^M - Y^H) + N_{\mathrm{R2}}(Y^M - 2Y^H)] - C_H$. Therefore, profit maximization with respect to P_H^{Trade} gives us $P_H^{\mathrm{Trade}} = (N_{\mathrm{R1}} + N_{\mathrm{R2}})2Y_M(Q_h^H - Q_h^L)/(N_{\mathrm{R1}} + 2N_{\mathrm{R2}})3Q_h^H$. Comparing the above expression with Eq. (11) we get $P_H^{\mathrm{Trade}} = \{(N_{\mathrm{R1}} + N_{\mathrm{R2}})/(N_{\mathrm{R1}} + 2N_{\mathrm{R2}})\}P_H$. Hence, we can say that $P_H > P_H^{\mathrm{Trade}}$, as $(N_{\mathrm{R1}} + N_{\mathrm{R2}})/(N_{\mathrm{R1}} + 2N_{\mathrm{R2}}) < 1$.

hypothetical partial equilibrium framework, where we assume that the high quality health services is provided by a Multinational Health Service Provider with two faces. From this model we have shown that at autarky the demand for high quality health services in the South will go up due to any kind of innovation or positive external shock to the existing high quality health care. Apart from this, in the present study we have shown that as a result of inflow of patients from the North to the South there is a fall in the cost of health quality signalling. Moreover, we have shown that trade in health service not only reduces the cost of health quality signalling but also improves the overall quality of health service in the South due to an increase in its demand. The demand for high quality health services in the South increases not only for an increase in the number of foreign patients but also because of an increase in the number of domestic patients who want to consume such high quality health services.

References

Acharyya R (2000) Income distribution, monopoly and quality distortion. Jadavpur University, Mimeo

Bagwell K, Riordan M (1991) High and declining prices signal product quality. Am Econ Rev 81:224–239

Banik A, De P, Bhaumik PK, Huq SMM, Bajracharya BB (2010) Trade in health services in South Asia: an examination of the need for regional cooperation. project report submitted on 10th round regional research competition of South Asia network of economic research institutes

Bhagwati JN (1984) Why are services cheaper in the poor countries? Econ J 94(374):279–286

Findlay C, Warren T (2000) Impediments to trade in services: measurement and policy implications. Routledge, New York

Garcia-Alanso MDC, O'Donnell O (2001) Income redistribution and access to innovations in health care. MPRA Working Paper Series

Goddeeris JH (1984a) Insurance and incentives for innovation in medical care. South Econ J 51:530–539

Goddeeris JH (1984b) Medical insurance, technological change and welfare. Econ Enquiry 22:56–67

Kelegama S (2009) Trade in services in South Asia. SAGE, India

Menck KW (2005) Medical tourism−a new market for developing countries. Travel Daily News. http://www.traveldailynews.com

Mudu G (2004) Hospitals in India woo foreign patients. Br Med J 328(7452):1338

Sampson GP, Snape RH (1985) Identifying the issues in trade in services. World Econ 8:171–181

Spence AM (1974) Job market signaling. Q J Econ 87(3):355–374

Tirole J (1988) The theory of industrial organization. MIT Press, Cambridge

Turner L (2007) First world health care at third world prices: globalization, bioethics and medical tourism. BioSocieties 2:303–325

Wolinsky A (1983) Prices as signals of product quality. Rev Econ Stud 50:657–658

Smuggling and Trafficking of Workers: A Brief Review and Analysis of the Economics of Illegal Migration

Saibal Kar

Abstract Studies on illegal labor migration to rich countries are strongly policy driven and welfare-centric. Border control and employer interdiction are the most popular policies for controlling illegal entries. We review a large body of literature on illegal immigration in general and smuggling and trafficking of workers in particular. A policy to lower exploitation from illegal migration is discussed with the help of an analytical note.

Keywords Illegal immigration · Migrant smugglers · Traffickers · Unemployment benefit · South Asia

JEL Classifications H21 · J33 · J48 · J61 · O15

1 Introduction

The issue of illegal immigration receives considerable attention from governments all over the world. The policy makers and the academia alike devote substantial interest in the subject owing to various sociopolitical and economic bearings on the recipient and the source countries. Defined broadly, illegal immigration constitutes a move from one country to another by ways and means not admissible under the general or country-specific legal requirements that 'regular' migrants need to satisfy. In some cases, however, overstaying by legal migrants turns them into illegal migrants and extends the definition of 'irregular migration.' However, not all developed countries are exposed to illegal immigration of serious magnitude. Statistically speaking, the United States is undoubtedly the single largest recipient

S. Kar (✉)
Centre for Studies in Social Sciences, Calcutta, R-1, B. P. Township,
Kolkata 700094, India
e-mail: saibal@cssscal.org

S. Kar
Institute for the Study of Labor (IZA), Bonn, Germany

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_14

of illegal immigrants; Mexico being the largest source country. The official statistics of the US Homeland Security Department states that there has been entry of an estimated 11.5 million people in USA in the year 2011 alone. When such statistics is reviewed globally by inclusion of a number of destination countries in Europe and Asia-Pacific (mainly Australia), the numbers are overwhelming. Not surprisingly, therefore, the staggering volume of illegal immigration has motivated many influential studies in economics, political science, sociology, and international law. These contributions are steadily growing and it may be useful to take a stock of the available issues in order to highlight some of the important policy implications emanating from these studies. To this end, we mainly offer a brief review of the literature on the economics of illegal migration with anecdotal reference to sociological and legal scholarship. As a caveat, we must also mention that even for the economics literature our purpose is not to offer a census of all studies, but review a few interesting results and policies.

The motivation behind reviewing this literature and discussing a brief policy-based descriptive model in the end is many. First of all, as we shall delineate with some data at the cross-country level, the cumulative magnitude of illegal immigration is quite high. In fact, the waves of relatively unplanned (host country perspective) and largely illegal immigration probably dating back to the rather famous incident of the Mariel Boatlift between Cuba and USA in the 1980s had always intrigued the researchers about what to expect on the labor market outcomes of such mass movements. In particular, the distinction of such labor mobility with regular migration principally has to do with the potential displacement and wage implications for the unskilled workers in the destination country. In this regard, Card (1990) finds that the Mariel immigrants increased the Miami labor force by 7 % and more than proportionately for the less-skilled occupations and industries. However, unlike in most other cases, the Mariel influx seems to have had virtually no effect on the wages or unemployment rates of less-skilled workers, even among Cubans who had immigrated earlier. The usual argument stands that such undocumented migrants easily become part of the ethnic and immigrant enclaves with the impact largely contained within the boundaries of such arrangements. In some cases, the spillover on to the larger economy is also observable. The author suggests that the ability of Miami's labor market to rapidly absorb the Mariel immigrants was largely owing to its adjustment to other large waves of immigrants in the two decades before the Mariel Boatlift. Not surprisingly, this is not the case always especially when the carrying capacity of all locations are rarely same, even though the location choice by skilled and unskilled, legal, or illegal immigrants is overwhelmingly in favor of larger cities. Apart from this argument, the other important issue that becomes way more complicated with waves of illegal migration is that of the provision of public goods at destination. While the specific interactions are discussed at length, for motivational purpose one can briefly argue that the allocation and share of public goods becomes an aspect of moral hazard both on the part of the provider and the co-beneficiaries. In fact, the well-known conversation between Mrs. Obama visiting a neighborhood school and a child from the primary section born of parents with illegal immigrant status, has been at the helm of such tension. Finally, a world continuously shaken by acts of religious and political fundamentalism makes illegal immigration a much more complicated subject worthy of consideration in the fields of labor and international factor movements, than it ever were three decades earlier.

Before we proceed to further details, let us invoke a more recent concern about the categorization of illegal immigration into two possible strands dominating the current *modus operandi*. These are, namely, 'smuggling' and 'trafficking' of illegal workers mainly from poor to rich countries and we shall devote special attention to this distinction in later sections. Note that, albeit movement of workers across various countries have always been a matter of concern, but the magnanimity of what is now deemed as a severe problem is consequent on the impact of illegal movements spilling over to the economic, social, and political lives of the legal residents at the destination and source countries. It came under the political and economic scanners of national governments and policymakers especially in countries that have historically relied on migrant workforce and continued to be migrant heaven even 30 years back. Much of the current impact assessments, therefore, treat illegal immigration as a host country problem.

All of the above taken together, policies to restrict illegal immigration have been primarily driven by the political and economic judgments at the destination capturing only a partial demand side view of the problem. One has to nevertheless appreciate the fact that since illegal immigration is principally governed by the conditions of excess supply of relatively unskilled workers, the socioeconomic factors at the source must also be reviewed with more alacrity. In other words, unless sufficient attention is devoted to crucial incentives and disincentives governing decisions to migrate illegally from the source countries, many restrictive policies designed at the destinations are not likely to yield adequate results or lessen the degree of exploitation that is intrinsic to illegal labor mobility. Moreover, an attempt to understand the supply side of the problem is also important in light of the fact that the richer destinations find it increasingly challenging to monitor and control illegal entries despite elaborate mechanisms in place (viz., Bandopadhyay and Bandyopadhyay 1998). As part of casual observations it is agreed that illegal immigration as well as overstaying has increased as legalized access to the richer countries have steadily dwindled over time. Even with the Mode 4 type of arrangements under the GATS, which facilitates entry of manufacturing and construction workers in richer countries, a potential applicant is required to satisfy several stringent criteria to qualify as a legal migrant. Besides, many of these schemes typically operate on a favored nation basis and are far from transparent. This is regularly discussed in the media (BBC 2008; CNN 2008, for example) and considered as an important factor that positively affects the scale of illegal immigration. Notwithstanding such temporary shocks generated by global business cycles and policies, various reports of the United Nations estimate the number of illegal entries at 4 million every year. This, as we have alluded earlier, constitutes both smuggled and trafficked workers. In this regard, the United Nations Office of Drugs and Crime (UNODC) states that smuggling of workers is the fastest growing activity in the world with profit currently estimated at $31.6 billion.

The rest of the paper is divided into four sections. Section 2 broadly explores the economics of illegal immigration, where theory and examples relevant to the North American immigration problem feature as dominant sources of information. Section 3 discusses the economics of smuggling vis-à-vis trafficking and looks into the source country perspectives in detail. Section 4 deals with some sociopolitical impacts of smuggling and trafficking and Sect. 5 concludes.

2 The Economics of Illegal Migration

Since 1990s the US immigration policies have focused on border enforcement rather than other forms of interior controls, especially on the US–Mexican borders despite the fact it has so far not been very successful in deterring illegal entries.[1] In the year 1997, the INS (currently, US Citizenship and Immigration Service) had developed a surveillance system in the form of Integrated Surveillance and Intelligence System (ISIS), which failed on many counts despite incurring huge costs on behalf of the US government (Koslowski 2011). It is well-known that the dominant policy choices in the host country are restricted to border controls and/or internal monitoring. The trade-off between the two has been first discussed in Ethier (1986). It suggests that a policy mix of border enforcement and internal worksite monitoring affects the illegal wage and the capture rate of illegal entrants in a complex manner. When the border enforcement (expressed in terms of resources devoted to this activity) is zero but the domestic enforcement (expressed in terms of the domestic output spent for this purpose) is at its maximum, the wage received by an illegal worker is equal to that they receive at home (being captured with a higher probability and deported). On the other hand, if the domestic enforcement is zero and all resource is spent on border enforcement, then those illegal workers who manage to cross the border earn as much as the legal unskilled workers at destination (assuming imperfect identification of legal and illegal workers by employers). The substitution between border enforcement and domestic enforcement yields a strictly convex to the origin relationship and raises the illegal wage as the country moves toward zero domestic enforcement. The task before a social planner is to obtain a combination of the two policies, such that the total cost of enforcement is minimized, the national welfare maximized and the policy target of a certain level of unskilled legal wage is achieved. According to this paper, if employer penalty for hiring illegal workers turns out to be a real social cost (lost output), then the optimal combination of the policies is at point *C* (Fig. 1 from Ethier 1986, p. 64). The policy duo offers a convex combination denoted by the curve *AB* and the slope of this curve at minus unity is the optimal choice (Fig. 1). As the policymaker chooses

[1]It is recently reported, however, that the illegal Mexican entry into USA has dropped 16 % in 2008 due to deployment of 3000 more border patrol and the tightening job market in construction and services (CNN 2008).

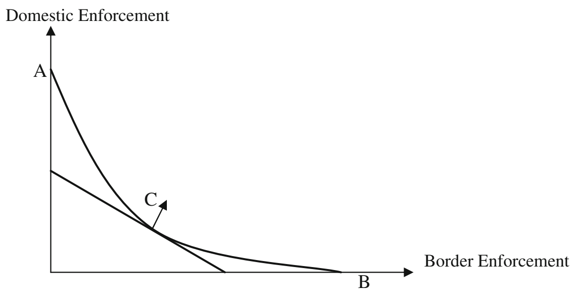


Fig. 1 Optimal choice between domestic and border enforcements

more border enforcement, the ratio of actual entries to potential attempts would fall, but so would detection at work. The illegal wage rises in the southeast direction. Movement toward the optimal point lowers net policy cost and improves *ex post* welfare of immigrants, but further movement in the southeast direction substitutes native welfare for illegal workers' welfare. Overall, the policy mix is effective in separating two distinct goals of income distribution and controlling of the volume of migration.

These findings influenced research on various dynamic implications associated with comparable policies aimed at controlling illegal immigration. Border enforcement, worksite inspection, and prohibition of employment of illegal workers all increase the entry cost for illegal workers. A large portion of the literature discusses how heightened border enforcement interacts with border crossing services—popularly called 'coyote' crossing in the US. Djajic (1999) used a simple dynamic framework of international and inter-sectoral migration of illegal workers to argue that tougher employer sanctions, intensive identity checks, apprehensions, and deportations help in lowering the stock of illegal immigrants only if their presence is limited to one geographic location and one economic activity. As tougher border controls raise the cost of illegal entry, illegal immigrants start investing in anti-detection efforts more. If the second effect dominates the first, the policy may have an adverse effect raising the stock of illegal immigrants.

Compared to this, employer sanctions are more effective in the sense that they reduce the flow of illegal immigrants by lowering their wages. Note that, an enhancement of identity checks, apprehensions, and deportations might become ineffective if the demand for agricultural labor is inelastic by nature. It has often been observed that illegal immigrants might venture into newer geographic locations through support networks in an attempt to evade detection by authorities. The study clearly suggests that deeper research into the networking abilities of illegal immigrants is required in order to suggest policies, which are complementary in nature and are capable of dealing with the ability of immigrants to pervade into areas that are less prone to detection by authorities.

Guzman et al. (2008) offer several conditions wherein the response of tighter border enforcement is countered by greater time investments in border crossing. However, if the 'smugglers' are more efficient and use capital-intensive techniques, less time is spent on border crossing attempts. This in turn raises the return from illegal migration and more workers choose to migrate illegally. In fact, many other possibilities exist, whereby the capital stock and saving propensity interact with the incentives to migrate illegally. Evaluation of the role of border enforcement vis-à-vis employer interdiction suggests that developed countries are unlikely to choose low border enforcement compared to no enforcement. Similarly, Woodland and Yoshida (2006) have discussed the connection between factor mobility and illegal immigration when illegal workers display risk aversion. What do these and similar policy propositions imply for welfare of the natives at destination?

2.1 Impact at Destination

The raging debate about illegal immigration and the distribution of wealth and economic development in sovereign countries is in stark contrast to the fact that people moved about quite freely even a century earlier.[2] However, the word 'passport' had come into existence at least by the sixteenth century when William Shakespeare uses it in Henry (1599). But, there could not have been as much restriction on where individuals would want to settle as long as it did not disturb the fair shares of the dominant groups and did not amount to *landnahme* (land-taking) by invaders.[3] The large intra-Europe migration clearly distinguished the Antiquity from the Middle Ages and spilled further on to the Medieval and Early Modern periods in Europe where Viking, Germanic, Slavic, Turkic, Avars, Bulgars, Angles, and Saxons moved about the entire continent. Modern day migration is characteristically different from Medieval and Early Modern migration in the sense that people do not move about in considerably large groups any longer. Decision to migrate is purely individualistic, albeit in some cases migrants use the social capital of groups in order to choose appropriate destinations and lower the cost of relocation. In contrast, illegal immigration is characterized by movements in groups (although, smaller) and therefore reminiscent of historical migration patterns.

The economic, social, and political impact of illegal immigration at the destination is looked at from various angles. First and foremost, many studies discuss implications of illegal immigration for the native job market via a number of questions, such as: (1) are the immigrants competing with native workers for the same jobs? (2) How are wages determined for illegal workers? (3) Is there a job market segmentation in effect? Most answers are critically linked to the illegal arrangements within the recipient countries and influenced by the mode that the illegal immigrants use for entry, whether smuggled in or trafficked. Generally, private firms or farms would always be inclined to hire workers willing to work for

[2]The only Indian to have arrived in USA in the year 1850, landed in Texas; US Bureau of Statistics.

[3]These included conquest of Pannonia by Hungarians, Indo-Aryan conquest, Franks arriving in Francia, Seljuk invasion of Anatolia, Gaels coming to Ireland, Anglo-Saxon settlements in Britain etc.

the lowest admissible wages. If the level of productivity at the lower rungs of the job ladder (where most of the illegal immigrants find jobs, see Djajic 1987) is not dependent on the legal status of a worker, firms are essentially indifferent between hiring a legal and an illegal worker. But, there are always legal and institutional restrictions on hiring illegal workers manifested in the form of penalties, revoking of license or legal sanctions on the errant firms, subsequently creating the segmentation between legal and illegal labor markets. The wages and conditions of work in each market are direct products of such segmentation. Carter (1999) suggests that the effects of illegal immigration may both benefit and hurt unskilled workers owing to targeting of primary sector jobs. When illegal immigration is low, availability of primary sector jobs is high and is shared by natives and illegal immigrants. However, as more illegal workers enter, more and more natives are displaced from primary jobs leading to Pareto-superior policy propositions in favor of apprehending and deporting illegal workers from the primary sector.

This hardly means, however, that the demand for illegal workers is zero. First, the act of monitoring firms is costly and requires political consensus in a democracy. And democracies such as USA, according to Bhagwati (1987, p. 32), "with their strong civil-libertarian and civil lobbies, are not in a position to start shooting happily at the illegal migrants".[4] The alternative therefore is to 'reason' with the firms not to hire illegal workers, in the same spirit as many developing country governments now campaign against hiring of child labor given the official ban on such employment. Undoubtedly, it leaves enough space for the firms to flout norms with a probability attached to being apprehended on the act and pay a penalty. If the firms mange to evade apprehension, they retain considerable rent generated by illegal immigration.

Note that, unwittingly enough, policies adopted by a country to curb illegal inflows often produce perverse outcomes for the legal labor market. The allure of amnesty (Karlson and Katz 2003) and a not-so-perfect border control leads to self-selection among illegal workers regarding who should migrate. It turns out that a positive probability of amnesty offered at the destination may attract better ability workers who will accept lower pays initially given the option to relocate to a higher paying job once the amnesty is granted. The low ability illegal workers on the other hand would not find (low) wage offers negotiable over time and will have little outside options given their skill—in addition to the looming threat of information leakage from the employing firms to the enforcement authorities. This would constitute a source of persistent exploitation of illegal workers as one observes from numerous sweatshop stories. It is further argued that unlike high ability counterparts, the low ability illegal immigrants become a burden to the host economy by consuming publicly supplied goods and services without paying taxes. A mix of amnesty and enforcement applicable to the illegal workers are likely to address the

[4] Originally published as Bhagwati (1979), International factor movements and national advantage, 9th Ramaswami Lecture, *Indian Economic Review* 14, 73–100.

inherent tension between these options, in the same spirit as the trade-off discussed in Ethier (1986).

In this connection, Thum (2004) considers how the public expenditures may indirectly turn into an instrument of migration control given that in some cases free mobility of labor may be difficult to impede, as currently true for the unified labor market in the European Union. Enactment of discriminatory practices with regard to public expenditure may be a viable policy when more direct practices such as charging higher taxes or excluding foreigners from public services are legally infeasible. More specifically, the government at the destination may predominantly provide those goods that benefit natives more than foreigners assuming that there is sufficient heterogeneity in language, culture, and the consumption baskets between natives and immigrants. Since immigrants (in this case legal, but equally applicable to illegal workers) are often only temporary guest workers, it should not be difficult for a government to isolate expenditure categories along preference patterns. For example, it is straightforward that irregular migrants are less likely to visit theaters, use little services from public libraries or consume long hours of public broadcasting services. However, if there are widespread social safety nets, recreational facilities, or public transport infrastructure, then immigrants cannot be fully excluded from these services, unless a law as the one recently adopted in Arizona comes into effect. It is argued that this law would potentially bring back the problems associated with racial profiling of immigrants in the US. According to this state law, all individuals must always carry identity proofs so that citizens could be distinguished from aliens. This is unusual for a country that did not use iron curtains for immigrants historically—legal and illegal. The proponents argue that this is a policy to achieve 'attrition through enforcement' in the number of illegal entries. In order to bypass this law, large number of immigrants of Hispanic origin has consequently moved over to New Mexico, the neighboring state. Nonetheless, this is reminiscent of the infamous segregation policies prevalent in the United States for a very long time and the threats facing Blacks, Hispanics, and other nonwhite immigrants in the form of political acceptability of the Ku Klux Klan (see, Levitt and Dubner 2009).

In other words, given the decision to migrate at all, the location choice is seriously influenced by public expenditure and enforcement characteristics in effect. In fact, Borjas (1999) shows that due to this differential location choice U.S. states with higher social transfers still become the welfare magnets for immigrants. These results, general for the immigrant population as a whole, are premised on the fact that immigrants do not have voting rights. Mayr (2007) derived conditions that if legal unskilled immigrants could vote then the tax rates could end up being too low hurting redistribution for all. Thus, out of pure re-distributive concerns alone the natives could vote against inflow of unskilled immigrants. This result does not entertain the possibility that such endogenous choices could exacerbate the problem of illegal unskilled immigration. These choices offer a basket of second best options.

The *ex post* theory of Epstein and Weiss (2001) that *ex ante* illegal immigrants are undesirable, is however, much less myopic. Since border patrol and internal monitoring both consume too much resource, it is not economically efficient to completely stem illegal inflow (also impossible in view of long and porous borders between USA and Mexico, in the same way as India and South Africa from time to time are exposed to illegal infiltration from politically unstable and economically underdeveloped neighboring countries). It is also possible that illegal immigration turns out to be an endogenous policy choice at the destination, wherein the illegal workers are deployed in sector-specific activities despite initial rejection of such migration policies by the median voter (Hillman and Weiss 1999). This reminds of the Ramaswami (1968) and Bhagwati (1979) propositions where it is ideal for a rich country, such as the US to allow all potential workers to move in from a poor country, like Mexico, and thereby marginal product of workers (and wage) lower at home.

The debate nonetheless continues in terms of the relationship between illegal immigration, minimum wages, and migration policies (see Vogel and Cyrus 2008 for similarities between the US and German policies; Gaytan-Fergoso and Lahiri 2004 for the relationship between foreign aid and illegal immigration, etc.). Generally, the willingness to work for low wages encourages owners of capital to employ illegal immigrants. This happens when the wages demanded are lower than the mandated minimum wage. However, when the employment of illegal immigrants increases, it exerts an upward pressure on the minimum wage. An immigrant would take up a job if the wage is higher than that offered in the source country. An increase in the minimum wage would decrease the number of legal workers employed. This helps to determine the relationship between the enforcement budget and the number of illegal workers employed. The enforcement budget is determined on the basis of the behavior of workers and employers. On the other hand, the effect of enforcement budget depends upon the relative strength of the (legal) labor union or capital owners union. When the labor union is stronger, an increase in minimum wage raises the optimal enforcement while the reverse occurs when the capital owners union is stronger (Epstein and Heizler 2007).

Accommodating further complications such as heterogeneous labor markets and growth in the presence of illegal migration Palivos and Chong (2010) show that illegal immigration increases average consumption levels of the economy and promotes skill acquisition among the natives. Further, it raises asset holdings of the skilled workers, but may also lead to unequal wealth distribution in the steady state as well as during the transition to a high illegal immigrant economy. It is possible that the transfer made from the accumulated penalty applicable to illegal immigrants (or employers) falls short of the loss endured by domestic unskilled workers. This leaves the wealth distribution even more skewed.

These and other arguments against illegal immigration often lead to persecution mania and misuse of the legal procedures aimed at reducing trafficking. This drives a considerably large section toward criminal activities and promotes growth of the shadow economies. On the other hand, the larger shadow economies prevailing in Asia, Latin America, and Africa keep creating opportunities for illegal immigration. Epstein and Weiss (2001) had suggested that illegal migrants might be considered for eventual amnesty to bring them back into the mainstream and thereby reduce additional costs incurred due to illegal 'activities' and deterioration of law and order situations. When one factor in the costs also borne by immigrants who choose to migrate illegally due to various reasons other than terrorism or similar intentions, the overall costs of illegal immigration becomes quite high. Chau (2001) argued that the constraint of offering amnesty to every illegal intruder works as a credible commitment on the part of the country to apprehend and deport illegal immigrants once they have entered the labor force. Section 3 deals with an analytical case where illegal immigration results from profit-seeking illegal activities by agents. It also brings up a set of complex interactions between migrant agents, local employers, source country issues, and therefore the role of public policies.

3 An Analytical Note with Smuggling and Trafficking of Illegal Workers

"Sonia was invited to come to the United States by family friends and told that she could work for them as a housekeeper, and they would pay her $100.00 a week. Sonia was provided with fraudulent documents and departed for the United States with her new employer. She knew that this was illegal, but she needed the money, and was willing to take the risk. *Was Sonia smuggled or trafficked? Sonia was smuggled into the United States.* She left willingly with full knowledge that she was entering the United States illegally.

Upon arriving in the United States, Sonia was kept in isolation, she was given a place to sleep in the basement and told not to speak to anyone or she would be turned over to the Immigration Service. Sonia was never paid for her work and felt that she had no one to turn to for help. *Was Sonia smuggled or trafficked?* At this point Sonia was restricted from leaving the house, threatened with deportation if she attempted to talk to anyone, and forced into involuntary servitude. *Sonia is a victim of trafficking.*"

(Excerpts from Fact Sheet, The Human Smuggling and Trafficking Center at the U.S. Department of State, Washington DC, January 1, 2005; italics added)

The excerpts clearly suggest that trafficking in person is not one-dimensional. There is a fine line between smuggling and trafficking and that statistically it may be quite difficult to isolate the two. In fact, what global data sets capture is essentially accounts of trafficked individuals who have been identified by public monitoring authorities and the information entered into official databases such as the European Central Aliens Register (US equivalent of this being the Central Index). Table 1 displays the top 10 destinations where trafficked migrants have been located in recent times.

According to UNODC statistics, there has been a rising trend in the number of victims of trafficking identified globally during 2003–2006. 71 countries feature in the list where the authorities have been able to identify the victims recording a rise in numbers from 11,706 in 2003 to 14,909 in 2006 (UNODC 2009). In addition, the profiles of victims were documented and aggregated in 61 countries for the year

Table 1 Top 10 countries of destination where victims were trafficked in 2011

Destination country	Numbers
Russian federation	837
Haiti	658
Yemen	552
Thailand	449
Kazakhstan	265
Afghanistan	170
Indonesia	148
Poland	122
Egypt	103
Turkey	101

Source IOM trafficking case data. This, however, includes cases of internal trafficking also

2006. It was observed that 66 % of the victims were women, 13 % were girls, 12 % were men, and 9 % were boys (UNODC 2009). Drbohlav et al. (2013) provide similar evidence on age profile of smuggled workers from Czech Republic.

The latest report (2012) of the U.S. State Department on the trafficking in persons provides further insights regarding prosecutions, victim identification, and convictions of traffickers from across the world. A cross-country assessment of the current scenario (based on estimates only, owing to lack of uniformity in national reporting structures) of trafficking in persons is provided in Table 2. This in fact has direct relations with the United Nations 3P, namely, Prosecution, Protection, and Prevention aimed at monitoring and containing trafficking in human beings. Of late, the demand side aspects of smuggling and trafficking have gained some importance in the policy forum globally. From South Asian perspective, however, the situation is quite grim. Table A.1 in Appendix 1 offers an index compiled from Cho et al. (2011) to report the status of this policy implementation for South Asian counties. It shows that between the year 2000 and 2009, India, Pakistan, Bangladesh, Nepal, and Sri Lanka, all have regularly prosecuted cases of trafficking and therefore score high on this count (on a scale of 1–5, meaning worst to best). However, except for a few years, protection offered to victims or potential victims and prevention in terms of concrete policy measures have been quite lackluster. Overall, therefore, the index for 3P is less than 3 for major South Asian countries, suggesting that there is room for substantial improvement (for detailed discussion on Asian migration patterns, see Abella 2003). There should be little doubt that preventing smuggling and trafficking of workers and lowering the level of economic cost and exploitation associated with it require a better understanding of how the policies might work in a complex maze of conflicting interests.

Thus, we offer an analytical discussion in this context where the 'second best' policy of creating disincentives for employers of illegal workers, for agents who become traffickers and the individual workers who consider migrating illegally, is explored. Related discussion is also available in Wheaton et al. (2010), which considers the market for traffickers as a monopolistically competitive one and

Table 2 Region-wise description of identification, prosecution, and conviction

Region	Year	Prosecutions	Convictions	Victims identified
South and Central Asia	2005	1041	406	
	2006	629	275	
	2007	824(162)	298(33)	
	2008	644(7)	342(7)	
	2009	1989(56)	1450(10)	3510
	2010	1460(196)	1068(11)	4357
	2011	974(24)	829(11)	3907
Africa	2005	194	58	
	2006	170	51	
	2007	123(28)	63(26)	5
	2008	109(18)	90(20)	7799
	2009	325(47)	117(30)	10861
	2010	272(168)	163(113)	9626
	2011	257(99)	218(116)	10094
East Asia and Pacific	2005	2580	2347	
	2006	1321	763	
	2007	1047(7)	651(7)	
	2008	1083(106)	643(35)	3374
	2009	357(113)	256(72)	5238
	2010	427(53)	177(9)	2597
	2011	1581(55)	1213(55)	5357
Europe	2005	2521	1792	
	2006	2950	1821	
	2007	2820(111)	1941(80)	
	2008	2808(83)	1721(16)	8981
	2009	2208(160)	1733(149)	14650
	2010	2803(47)	1850(38)	8548
	2011	3162(271)	1601(81)	10185
Western Hemisphere	2005	170	59	
	2006	443	63	
	2007	426(1)	113(1)	
	2008	448(42)	161(24)	6609
	2009	647(47)	553(66)	9020
	2010	732(80)	293(65)	6681
	2011	1023(42)	318(52)	9836
Near East	2005	112	104	
	2006	295	187	
	2007	415(181)	361(179)	
	2008	120(56)	26(2)	688
	2009	80(9)	57(8)	1011
	2010	323(63)	68(10)	1304
	2011	209(17)	60(5)	1831

Source Department of State of United States of America (2012)
Note The numbers in parentheses are those of labor trafficking prosecutions and convictions

analyze decisions made by potential migrants, employers of trafficked workers and social conditions that influence choice of trafficking by individuals. A number of other contributions are also available in the International Migration (2010). Existing studies on trafficking and exploitation emanate mainly from legal, political, and sociological documents (McCreight 2006; Granville 2004; Kyle and Koslowski 2001; Abella 2000; IOM 2000; Bales 1999; etc.), and only recently have motivated economic analysis. Tamura (2010) shows that migrant smugglers differ in exploitative powers (also see, Di Tommaso et al. 2009). Based on the definitions of *smugglers* (organization that provides illegal border crossing services) and *traffickers* (organization that provides illegal border crossing services but exploits its clients after smuggling, UN 2000a, b), it determines endogenous distribution of smugglers and traffickers. Policy choices include border enforcement and high penalty for smugglers/traffickers. The critical policy implications are that border enforcement reduces smuggling activity whereas, improved inland monitoring *only* might induce non-smugglers to take up smuggling.

We use a similar source–destination analytical structure involving migration (also see, Woodland and Yoshida 2006), for isolating smugglers from traffickers. Instead of beginning with the usual policies adopted for curbing illegal immigration, we consider a standard tax-unemployment benefit approach. Not surprisingly, empirical evidence connecting illegal immigration to tax and unemployment benefit is unavailable.

Let us consider a large illegal system where migrant smugglers, traffickers, local employers, and potential migrants, all respond to a policy choice made in the destination country. However, before we proceed further on this, let us once again define the scope of the two types of migration that form the basis of this model. Trafficking in Persons means “the recruitment; transportation; transfer; harboring or receipt of persons; by means of threat or use of force or other forms of coercion; of abduction; of fraud; of deception; of abuse of power or position of vulnerability or of giving or receiving payments or benefits to achieve the consent of a person having control over another person; for the purpose of exploitation.” On the other hand, the Smuggling of Migrants Protocol supplementing the United Nations Convention against Transnational Organized Crime defines the smuggling of migrants as the “procurement, in order to obtain, directly or indirectly, a financial or other material benefit, of the illegal entry of a person into a State Party of which the person is not a national or a permanent resident” (Article 3, Smuggling of Migrants Protocol, UNODC).

Given this distinction we assume that smugglers and traffickers function as intermediate agents and their distribution is endogenous. The government’s objective function in the context of illegal immigration represents that of legal unskilled workers in the recipient country.[5] Workers pay an optimal tax according to this objective function for financing inland monitoring, at a given level of border

[5]Skilled workers often vote against illegal immigration since it affects general conditions of living, increases crime and dependence on transfer payments go up. However, their jobs are not directly threatened by it. Presently, we leave out unemployment concerns of skilled workers. Further, the government’s objective function may directly include concern for illegal immigrants. We show

enforcement (see further, Singer and Massey 1997; Hanson et al. 1999; Bandopadhyay 2006). Smugglers and traffickers are distinguished in terms of the exploitative rent they extract from illegal immigrants.

The utility function of an unskilled native or legal immigrant is our basic premise. We assume that it represents the social utility function as far as illegal immigration is concerned. Choice of an optimal tax required to fund inland monitoring against illegal workers (given an exogenous level of unemployment benefit declared by the government) determines the level of penalty that should be imposed on local employers and traffickers. Note that, both local employers and traffickers hire illegal workers, but there are subtle differences. The local employer is either informed or not informed about the resident status of the worker—unless racial profiling is a dominant characteristic of such equilibrium—but the trafficker is certainly aware of it. Now, the determination of optimal tax borne by the legal workers and the levy of penalty on employers of illegal workers together should determine the market-clearing wage for illegal workers. Using these, one obtains the payoffs of traffickers and smugglers. A migrant smuggler or 'coyote' only aids in border crossing services and leaves the illegal migrant to find support on his own (usually in local farms). Traffickers, as already defined, control the activities of smuggled individuals, keep them in captivity and extract rents above the fee for border crossing. Acting as a smuggler exposes the agent to border control only, whereas a trafficker faces internal interdiction in addition to border control. If successful, the additional risk pays off as higher rent extracted from the trafficked migrant who are used as forced unpaid labor. At the point where the returns from the act of smuggling and trafficking are identical, an agent is indifferent between the two. Those who command greater power of rent extraction end up being traffickers and others distributed with rent extracting power lower than the critical level become smugglers.

On the supply side of the model the prospective immigrant from the developing country is asymmetrically informed about true identity and exploitative power of smugglers and traffickers (the excerpt in Sect. 3 refers to this characteristic). They make decisions on whether or not to migrate depending on expected share of smugglers and traffickers. The expectation is an important criterion for determining expected income of a potential illegal migrant in the foreign country.

Given this intuitive setup, we consider a policy. Assume an exogenous increase in the rate of unemployment benefit. As the government finances inland monitoring via resources generated from penalty (with a positive probability perpetrators are captured, and otherwise), a rise in committed unemployment benefit may only be covered by a higher penalty imposed, given the tax revenue already collected. This, in turn must lower illegal wage. Since the employers of illegal workers now face higher penalty, they transfer the burden to the illegal workers in terms of lower wage. A lower wage should dissuade illegal migrants and is also expected to lower

(Footnote 5 continued)

that the use of tax and unemployment benefit is a means to tackle illegal immigration even if the objective function does not explicitly include such arguments.

profits of both smugglers and traffickers both of which extract portions of the wage either as commission or as manifestation of exploitative power. However, agents who command a high exploitative power shall remain in business and the rest will either turn to smuggling or go out of business. A lower expected wage and a higher possibility of matching up with traffickers, lowers expected foreign income of the prospective illegal worker and it should desist inclinations to migrate illegally. Thus, it is possible to lower exploitation from illegal migration with a standard labor market policy usually in practice in most destination countries.

It should be clearly pointed out here that we do not expect the usual productivity and participation-related effects of higher unemployment benefit to arise in this case. To elaborate, one may find analogy with the well-known wage subsidy policy in vogue in the US. The trade adjustment assistance allows unemployment benefit to flow to a worker *only* if the worker can establish that the loss of job has been exclusively owing to penetration of cheap foreign commodities or outsourcing of production to a cheaper location. The relevant literature is abound with such examples. Similarly, the unemployment benefit for such unskilled workers as may be displaced by lower wage illegal workers has very little impact on effort or productivity at the economy-wide level. Furthermore, the usual onus is on the displaced worker to establish such connection and therefore the equilibrium configuration of this strategy seems stable.

In poor countries, the prospect of illegal migration often comes at a very high cost, including sale of land and other assets in order to finance illegal border crossing. Lower foreign expected income compared to the high cost borne by a potential migrant is likely to create disincentives for migration. Of course, an exogenous rise in the unemployment benefit may not meet with success if looked at in isolation or from the perspective of the entire labor market. It is the budget constraint of the recipient country that dictates the optimal penalty imposed, which in turn is directly proportional to the commitment on unemployment benefit. The higher unemployment benefit and the higher penalty create dual pressure on the employers of illegal workers. Overall, under plausible conditions, the policy may turn out to be welfare improving at both ends.

4 Concluding Remarks

In the entire literature, standard deterrent policies involve the perpetrators, the direct victims and the government representing interests of various sections of the population—the capitalists, large farmers, skilled workers, etc. Policies to control illegal immigration run into complicated domains of lobbying, human rights or bilateral treaties between countries. In related studies general well being of trafficked and exploited workers, creation and sustenance of illegal immigration and bonded labor, specific implications of minimum wage on illegal immigration, ban on prostitution, migration reforms and amnesty and debt contract and persistence of exploitation have been discussed.

This review and the brief analytical discussion need substantially more empirical observations to help concrete policy prescriptions. According to Antonio Mario Costa (UNODC), "what counts mostly is the exploitation that takes place at several points along the chain as the human trafficking takes place and that is repetitive and prolonged" (BBC 2008). Many UN member countries have not yet ratified the Anti-Trafficking Protocol, and not much is known about the scale of activities. The efforts from academics, politicians and policy makers nevertheless offer a rich set of cases and policies, which eventually must lead to better outcomes than the second best policies discussed all through this review.

Acknowledgments The author thanks an anonymous reviewer for useful comments. The usual disclaimer applies.

Appendix 1

See Table A.1.

Table A.1 Country index and achievements for United Nations 3P

Code	Country	Year	Prosecution	Protection	Prevention	Overall 3P
AFG	Afghanistan	2000				
AFG	Afghanistan	2001	1	1	1	3
AFG	Afghanistan	2002				
AFG	Afghanistan	2003	2	3	3	8
AFG	Afghanistan	2004	3	3	3	9
AFG	Afghanistan	2005	3	2	3	8
AFG	Afghanistan	2006	2	2	3	7
AFG	Afghanistan	2007	2	2	2	6
AFG	Afghanistan	2008	2	2	2	6
AFG	Afghanistan	2009	2	2	3	7
BGD	Bangladesh	2000	4	3	3	10
BGD	Bangladesh	2001	4	3	3	10
BGD	Bangladesh	2002	5	3	4	12
BGD	Bangladesh	2003	5	2	3	10
BGD	Bangladesh	2004	5	4	3	12
BGD	Bangladesh	2005	5	3	4	12
BGD	Bangladesh	2006	5	3	3	11
BGD	Bangladesh	2007	5	3	3	11
BGD	Bangladesh	2008	5	3	3	11
BGD	Bangladesh	2009	5	3	3	11
IND	India	2000	4	2	2	8
IND	India	2001	4	3	3	10

(continued)

Table A.1 (continued)

Code	Country	Year	Prosecution	Protection	Prevention	Overall 3P
IND	India	2002	4	3	3	10
IND	India	2003	5	2	3	10
IND	India	2004	5	3	4	12
IND	India	2005	4	3	2	9
IND	India	2006	5	3	2	10
IND	India	2007	4	2	2	8
IND	India	2008	4	2	3	9
IND	India	2009	4	2	3	9
NPL	Nepal	2000	4	3	3	10
NPL	Nepal	2001	4	4	3	11
NPL	Nepal	2002	4	3	3	10
NPL	Nepal	2003	4	3	3	10
NPL	Nepal	2004	5	3	4	12
NPL	Nepal	2005	5	3	3	11
NPL	Nepal	2006	5	3	3	11
NPL	Nepal	2007	4	2	3	9
NPL	Nepal	2008	4	2	3	9
NPL	Nepal	2009	4	2	3	9
PAK	Pakistan	2000	2	2	1	5
PAK	Pakistan	2001	3	2	3	8
PAK	Pakistan	2002	4	2	3	9
PAK	Pakistan	2003	4	2	2	8
PAK	Pakistan	2004	5	4	4	13
PAK	Pakistan	2005	5	2	4	11
PAK	Pakistan	2006	5	2	3	10
PAK	Pakistan	2007	5	2	3	10
PAK	Pakistan	2008	4	2	2	8
PAK	Pakistan	2009	4	2	4	10
LKA	Sri Lanka	2000	4	3	3	10
LKA	Sri Lanka	2001	4	2	3	9
LKA	Sri Lanka	2002	4	3	3	10
LKA	Sri Lanka	2003	4	3	3	10
LKA	Sri Lanka	2004	4	3	3	10
LKA	Sri Lanka	2005	4	2	3	9
LKA	Sri Lanka	2006	4	2	3	9
LKA	Sri Lanka	2007	4	2	3	9
LKA	Sri Lanka	2008	4	2	3	9
LKA	Sri Lanka	2009	4	2	3	9

Source Cho et al. (2011)

Note Prosecution Score 1 (worst) to Score 5 (best); *Protection* Score 1 (worst) to Score 5 (best) *Prevention* Score 1 (worst) to Score 5 (best). *Overall Index* Overall 3P. Score 3 (worst) to Score 15 (best)

References

Abella M (2003) Driving forces in Asian migration, in world migration report 2003. International Organization for Migration, Geneva
Abella M (2000) Migration and employment of undocumented workers: do sanctions and amnesties work? In: Cinar D, Gächter A, Waldrauch H (eds) Irregular migration: dynamics. Impact, Policy Options, Vienna, European Centre Vienna, pp 205–215
Bales K (1999) Disposable People: New Slavery in the Global Economy, Berkeley: University of California Press
Bandyopadhyay S (2006) Illegal immigration and second best import tariffs. Rev Int Econ 14 (1):93–103
Bandyopadhyay S, Bandyopadhyay SC (1998) Illegal immigration—a supply side analysis. J Dev Econ 57(2):343–360
Bhagwati, Jagdish N (1979) International factor movements and national advantage. Indian Econ Rev 14:73–100
Bhagwati, JN (1987) On the choice between capital and labor mobility. In: Feenstra R (ed) Essays in International Economic Theory by Bhagwati JN, vol. 2. MIT Press, Cambridge, pp 31–43
Borjas GJ (1999) Immigration and welfare magnets. J Lab Econ 17:607–637
British Broadcasting Company, BBC (2008) Trafficking: a very modern slavery, BBC News Report by Stephanie Holmes (February 15)
Card David (1990) The impact of the Mariel Boatlift on the Miami labor market. Ind Labor Relat Rev 43(2):245–257
Carter T (1999) Illegal immigration in an efficiency wage model. J Int Econ 49:385–401
Chau N (2001) Strategic amnesty and credible immigration reform. J Labor Econ 19(3):604–634
Cho, S-Y, Dreher A, Neumayer E (2011) The spread of anti-trafficking policies—evidence from a new index, CEGE Discussion Paper 119, March
CNN (2008), Immigration economics: Lou Dobbs reporting, March 21, 2008
Department of State of United States of America. (2012). Trafficking In Persons Report, Washington DC
Di Tommaso M, Shima I, Strøm S, Bettio F (2009) As bad as it gets: well-being deprivation of sexually exploited trafficked women. Eur J Polit Econ 25(2):143–162
Djajić S (1999) Dynamics of immigration control. J Popul. Econ 12:45–61
Djajic S (1987) Illegal aliens, unemployment and immigration. J Dev Econ 25:235–249
Drbohlav D, Štych P, Dzúrová D (2013), Smuggled versus not smuggled across the Czech border. Int Migr Rev 47(1):207–238, Spring
Epstein G, Heizler O (2007) Illegal migration, enforcement and minimum wage, IZA discussion paper # 2830, Bonn
Epstein G, Weiss A (2001) A theory of immigration amnesties, CEPR discussion paper no. 2830, London
Ethier WJ (1986) Illegal immigration: the host-country problem. Am Econ Rev 76(1):56–71
Gaytan-Fregoso H, Lahiri S (2004) Controlling Illegal Immigration: Border Control vs Domest Enforcement. J Quant Econ New Ser Spec Issue 2(2):203–215
Granville J (2004) From Russia without love: the "fourth wave" of global human trafficking. Demokratizatsiya, Winter
Guzman M, Haslag J, Orrenius P (2008) Coyote crossings: the role of smugglers in illegal immigration and border enforcement, Federal Reserve Bank of Dallas, research department working paper 0201
Hanson G, Robertson R, Spilimbergo A (1999) Does border enforcement protect U.S. workers from illegal immigration? National bureau of economic research working paper no. 7054, Cambridge: Mass
Hillman A , Weiss A (1999) A theory of permissible illegal immigration. Eur J Polit Econ 15:585–604

International Organization of Migration (2010) International migration. Spec Issue Spec Issue Hum Traffick 48(4):1–200
International Organization for Migration (2000) Migrant trafficking and human smuggling in Europe: a review of the evidence with case studies from Hungary. International Organization for Migration, Poland and Ukraine, Geneva
Karlson S, Katz E (2003) A positive theory of immigration amnesties. Econ Lett 78(2):231–239
Koslowski R (2011) The evolution of border controls as a mechanism to prevent Illegal Immigration. Report, Migration Policy Initiatives, February, Washington D C
Kyle D, Koslowaski R (eds)(2001) Global human smuggling: comparative perspectives. Johns Hopkins University Press, Baltimore
Levitt S, Dubner S (2009) Freakonomics: A rogue economist explores the hidden side of everything. William Morrow Paperbacks, NY
Mayr K (2007) Immigration and income redistribution: a political economy analysis. Public Choice 131(1):101–116
McCreight M (2006) Smuggling of migrants, trafficking in human beings and irregular migration on a comparative perspective. Eur Law J 12(1):106–129
Palivos T, Chong K (2010) Illegal immigration in a heterogeneous labor market. J Econ 101:21–47
Ramaswami V.K (1968) International factor movement and the national advantage. Econ 35:309–310
Singer A, Massey D (1997) The social process of undocumented border crossing, UCLA Lewis Center for Regional Policy Studies Working Paper No. 27, Los Angeles
Tamura Y (2010) Migrant smuggling, J Public Econ 94:540–548
Thum M (2004) Controlling migration in an open labor market. Public Choice, Springer 119(3–4):425–443
United Nations (2000a) Protocol to prevent, suppress and punish trafficking in persons, especially women and children, supplementing the United Nations convention against transnational organized crime, New York
United Nations (2000b) Protocol against the smuggling of migrants by land, sea and air, supplementing the United Nations convention against transnational organized crime, New York
UNODC (2009) Global Reporting on Trafficking in Persons, UN: Vienna
US Department of State (2012), Trafficking in Persons Report 2012
Vogel D, Cyrus N (2008) Irregular migration in Europe—doubts about the effectiveness of control strategies. Focus Migration Policy Brief, HWWI
Woodland A, Yoshida C (2006) Risk preference, immigration policy and illegal immigration. J Dev Econ 81(2):500–513
Wheaton E, Schauer E, Galli T (2010) Economics of human trafficking. Int Migr Spec Issue Spec Issue Hum Traffic 48(4):114–141

Impact of Trade Restriction on Child Labour Supply and the Role of Parents' Utility Function: A Two Sector General Equilibrium Analysis

Biswajit Chatterjee and Runa Ray

Abstract The use of child labour has been widespread across developing nations. Although the incidence of child labour use has declined in recent years in the aggregate, its use has remained been quite widespread in different developing nations. Various policy interventions have been suggested and debated in international bodies to combat or reduce the incidence of child labour use in different activities in poor labour abundant countries. The present paper develops a general equilibrium framework consisting of two sectors and three factors of production to assess the efficacy of restrictive trade policies on the incidence of child labour. The paper, derives the supply function of child labour using different kinds of utility function of adult worker (parent), with the interesting result that trade restrictive policy may fail to reduce the incidence of child labour use. This result challenges the popular view in favour of imposing trade sanctions on the import of those goods from the developed countries which are produced using child labour.

Keywords Child labour · General equilibrium

JEL Classification Numbers F10 · J13

1 Introduction

Child labour has been widely used in the developing world, although its incidence has diminished slightly in recent years, yet its use has remained quite high at the global scale. Policies have been suggested to combat the use of child labour in both developed and developing countries of the globe, yet the effectiveness of such

B. Chatterjee (✉)
Jadavpur University, Calcutta 700032, India
e-mail: chatterjeeb@vsnl.net

R. Ray
Vidyasagar College, Calcutta 700006, India
e-mail: runa_maju@yahoo.co.in

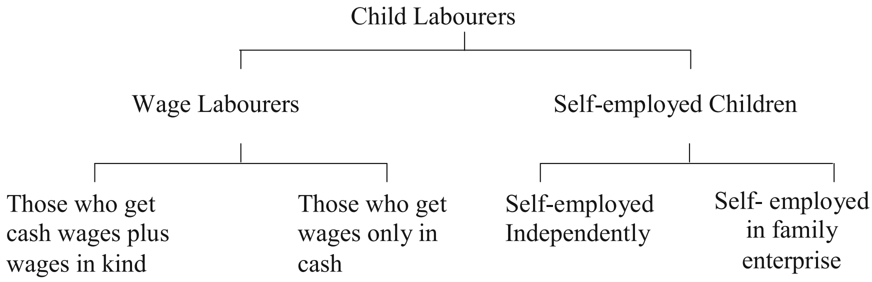
© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_15

measures have remained limited. Such measures include banning the use of child labour through legislations by the states and various support schemes to help the poor children to come out of work and go to school. But the incidence of child labour has remained quite high. According to the estimates by the ILO (2014), 168 million children worldwide are in child labour, accounting for almost 11 % of the child population as a whole. Children in hazardous work that directly endangers their health, safety and moral development make up more than half of all child labourers, numbering 85 million in absolute terms. The largest absolute number of child labourers is found in the Asia and the Pacific region but sub-Saharan Africa continues to be the region with the highest incidence of child labour. According to the report, the decline in child labour was greatest during the most recent 4-year period (2008–2012). The Asia and the Pacific region registered by far the largest absolute decline in child labour among 5–17 year-olds for the 2008–2012, and there had been reduction of the use of child labour by about one-third during the period 2000–2012, of which there was a reduction of 40 % in the number of girls in child labour as compared to 25 % for boys.

Following the Uruguay Round (UR) Agreement at Maracas, the World Trade Organization (WTO) was formed with effect from January 1, 1995, with a view to ensure fair and free trade practices among member countries. The WTO has brought to focus the issues relating to labour standards and environment. Regarding labour standards, the developed countries highlighted two points: the first relates to core labour standards and the second to child labour. Regarding the use of child labour, the demand of the developed nations has been that trade in commodities using child labour in many developing nations is unfair, and be banned—the children should go to school rather than to work. This position sounds fair on ethical and normative grounds, but it ignores the hard realities of the poor developing countries, where child labour use is conditioned by economic compulsions of the poor parents. Children in these countries are sent to work because their families are too poor to send them to school. When countries in the European Union boycotted the export of garments manufactured in Bangladesh on the ground that child labour was being used in manufacturing them, thousands of children lost jobs and landed up in other worse paid jobs. The fact is that developing countries are not in a position to take trade measures against them to eliminate the incidence of child labour employment.

The term ‘Child Labour’ is often used as a synonym for ‘employed child’ or ‘working child’. Although a precise definition of child labour is very difficult, it may be defined as that segment of the child population which participates in work, either paid or unpaid, working with the parents either inside family occupations or outside. Any work done by children that interferes with their full physical development, the opportunities for a desirable level of education and their needed recreation are called child labour (Stein and Davis 1940). We can distinguish alternative forms of child labourers visible in the following way:



The prevalence of child labour on grounds of economic necessities of the poor households in developing countries may be explained in terms of a demand–supply framework. In all the three major sub-sectors of such an underdeveloped economy—agriculture, manufacturing and services, there are demands for child labour, although the exact incidence varies across sectors. Generally, child labourers are employed in the agriculture sector, helping in agricultural operations like sowing, weeding, harvesting and threshing, and second, such child workers are also employed in sectors allied to fishing. Cottage and household industries constitute the third important sector within the secondary sub-sector in which child labour is concentrated. A number of factors explain the concentration of children in these sectors. First, such industries can easily evade legal restrictions. Second, it is believed that the human body is soft, flexible and pliable in childhood and therefore, and learn with ease how to twist, bend, stretch or assume different gestures as required by particular job. This is the 'nimble finger' argument. Moreover, children are more amenable to discipline and control—they can be coaxed, admonished, pulled up and punished for defaults without jeopardizing relations. Children are not organized and thus are paid a minimum wage to work for long hours. Finally, since the productivity per man in several such industries, mentioned above, is low, these enterprises cannot afford to hire adult workers, who are often organized and would demand higher wages. Thus, the principal motive of employment of child labour is to minimize costs to have comparative advantage in world market and thus to maximize profits, as many of these small scale enterprises produce for the world market. In all other sectors, the concentration of child workers is much below than that of general workers. The concentration is the least in the case of transport, storage and communication, banking, insurance (mostly services sector activities) followed by construction, mining. In urban areas, children are engaged in more diversified activities such as loading-unloading, hawking, looking after parked vehicles, rag-picking and shoe-shining. They are also employed as helpers in tea-shops and auto-repair shops or are engaged as domestic servants.

Trade sanctions, import tariffs and product labelling (for example, the Rugmark initiative in the carpet industry) has been proposed and in some instances implemented to reduce the extent of child labour. However, it has been recognized by several authors that such sanctions may not have their intended effect (see Bhagawati 1995; Dixit 2000; Jafarey and Lahiri 2002; Maskus 1997; Ranjan 2001).

Edmonds and Pavcnik (2002) find empirical support for the concern that trade sanctions may indeed raise child labour in the rice sector in Vietnam using data from 1993 to 1998.

On the supply side, child labour is often viewed as being driven by the needs of basic survival in developing nations, where children need to work to sustain themselves or their families. Basu and Van (1998) recognize this and establish that even through adults dislike child labour, they may have to endure it for survival. On the other hand, if the endowment/income level of family is much high, then these same households will not supply child labour. In other words, there are at least two reasonable equilibrium, one where children work and the other where they do not. Therefore, children work in developing nations not because of a lack of potential consciousness, but rather because of sheer economic necessity. Baland and Robinson (2000) also find altruistic parents enduring child labour because of poverty (the case of zero bequests) or imperfect capital markets. Higher family incomes, improved educational opportunities (Basu 1999; Maskus 1997; Ray 2002), and a general increase in the living standards of the poor through liberalized trade and improved access to world markets. (as in Bhagawati 1995; Dixit 2000) are expected to reduce child labour.

The neo-classical theory of child labour supply explains that the household or family supplies child labour for jobs in order to maximize its current income from the employment of the child labour in preference to the income expected from their employment in future after the schooling of its children. In fact, many of the households in a developing country decide for maximizing the present income rather than for the future income, which involves higher opportunity cost of schooling and investment of time and income going in for the future income. Moreover, the low life expectancy of children in developing countries is another reason why a household decides to maximize the present income rather than the future one.

In this paper, we consider a competitive *two* sector general equilibrium model of a small open less developed economy, which exports products using child labour. Child labour is specific to export sector and capital is specific to import-competing sector and adult labour is mobile between the sectors. Our objective in this paper is to examine the effectiveness of restrictive trade policies on the incidence of child labour. Rest of the world imposes trade restrictions on the exported product of the small open economy which is produced using child labour. In this paper, we have derived the supply function of child labour using different kinds of utility function of adult worker (parent), with the interesting result that trade restrictive policy may fail to reduce the incidence of child labour use. This result thus challenges the popular view in favour of imposing trade sanctions on the import of those goods from the developed countries which are produced using child labour.

2 The Model

The assumptions of the model are as follows:

(i) We consider a small open economy.

(ii) The economy consists of two sectors and three factors of production. Sector 1 produces an exportable commodity (X) with the help of adult labour and child labour. In this paper, we assume that rest of the world (ROW) imposes trade restrictions in such a way that exporters of the small open economy has to pay tax t on per unit export of X, resulting in reduction of the effective producer's price of X. Sector 2 is import-competing sector of the economy. It produces importable commodity (Y) with the help of adult labour and capital. Thus, adult labour in our model is perfectly mobile between the sectors but child labour and capital are specific factors of production. Child labour is specific it is specific in use in the export sector and capital is specific in import-competing sector of the economy.

(iii) We assume that adult labour is substitute for child labour. It is assumed that an adult labour is equivalent to β number of child worker where $\beta > 1$. Each adult worker earns a wage W^A. The child wage rate W^C must be (W^A/β) when adult wage rate is W^A.

(iv) Production functions exhibit constant returns to scale with diminishing marginal productivity to each factor.

(v) We assume that all inputs are fully employed and all markets are perfectly competitive. Stock of capital and adult labour force are exogenously given.

(vi) The prices of the traded goods X and Y are given internationally, due to our assumption of a small open economy.

(vii) Each firm maximizes profit.

(viii) In our model, supply of child labour originates from the household comparison of parental utilities from sending the child to work or to school.

The following symbols will be used in the formal presentation of the model:

$a_{L_a i}$: Adult labour-output ratio in the ith sector, $i = X, Y$

$a_{L_C X}$: Child labour-output ratio in the X sector

a_{KY}: Capital-output ratio in the Y sector

θ_{ij}: Distributive share of the ith input in the jth sector,
$i = L_a, L_C, K$ and $j = X, Y$

λ_{ij}: Proportion of the ith input employed in the jth sector,
$i = L_a, L_c, K$ and $j = X, Y$

P_i: World price of the ith good, $i = X, Y$

t: Ad-valorem tariff rate on the export of X

W^A: Adult wage rate

$W^c = \left(\frac{W^A}{\beta}\right)$: Child wage rate

r: Rate of return on capital

L: Adult labour endowment

L_C^S: Aggregate supply of child labour
K: Capital stock of the economy
$\wedge$: Proportional change
X: Domestic production of exportable
Y: Domestic production of importable

3 Derivation of Supply Function of Child Labour

In our model, the supply function of child labour is derived from the utility maximizing behaviour of the representative adult worker (parent). His utility function is given by

$$U = U(X, Y) - U(L_C^S) \tag{1}$$

The worker derives utility from the consumption of the final goods and disutility from child labour. For analytical simplicity, let us consider the following specific algebraic form of the utility function:

$$U = X^\alpha + Y^\alpha - \left(L_C^S\right)^\alpha \quad 0 < \alpha < 1 \tag{2}$$

It satisfies all the standard properties. Also it is additive and symmetric. It is homogeneous of degree α and has constant elasticity of substitution between any two arguments. The adult worker maximizes the utility function subject to the budget constraint.

In our model, we have not considered non-labour income. The budget constraint is given by

$$P_X(1-t)X + P_Y Y = \left[\left\{\left(\frac{W^A}{\beta}\right)L_C^S + W^A L + rK\right\}\Big/ L\right] \tag{3}$$

The Lagrangian expression for the constrained optimization problem is

$$M = \left[X^\alpha + Y^\alpha - \left(L_C^S\right)^\alpha\right] + \lambda\left[P_x(1-t)X + P_Y Y - \left\{\left(\frac{W^A}{\beta}\right)L_C^S + W^A L + rK\right\}\Big/ L\right] \tag{3a}$$

First-order conditions (FOCs) for constrained optimization problem are as follows:

$$\delta M/\delta X = \alpha X^{\alpha-1} + \lambda P_X(1-t) = 0 \tag{4}$$

$$\delta M/\delta Y = \alpha Y^{\alpha-1} + \lambda P_Y = 0 \tag{5}$$

$$\begin{aligned} &\delta M/\delta L_c^s = -\alpha (L_c^s)^{\alpha-1} - \lambda (W^A/\beta)/L = 0 \\ &\text{or,} \quad \alpha (L_c^s)^{\alpha-1} + \lambda (W^A/\beta)/L = 0 \end{aligned} \tag{6}$$

$$\delta M/\delta\lambda = P_X(1-t)X + P_Y Y - \left[\left\{\left(\frac{W^A}{\beta}\right)L_c^s + W^A L + rK\right\}/L\right] = 0 \tag{7}$$

$$\text{From FOCs we get,} \quad X = \left[P_X(1-t)/\left(\frac{W^A/\beta}{L}\right)\right]^{1/(\alpha-1)} L_c^s \tag{8}$$

$$Y = \left[P_Y/\left(\frac{W^A/\beta}{L}\right)\right]^{1/(\alpha-1)} L_C^S \tag{9}$$

Substituting the values of X and Y in the budget constraint (3) we get,

$$\begin{aligned} &P_X(1-t)\left[P_X(1-t)/\left(\frac{W_A/\beta}{L}\right)\right]^{1/(\alpha-1)} L_C^S + P_Y\left[P_Y/\left(\frac{W_A/\beta}{L}\right)\right]^{1/(\alpha-1)} L_C^S \\ &\quad = \left[\left\{\left(\frac{W_A}{\beta}\right)L_C^S + W^A L + rK\right\}/L\right] \end{aligned} \tag{10}$$

Simplifying (10) we get,

$$L_c^s = \frac{\frac{\left(W^A\right)^{\frac{\alpha}{\alpha-1}}}{(\beta L)^{\frac{1}{\alpha-1}}} + \left(\frac{W^A/\beta}{L}\right)^{\frac{1}{\alpha-1}}\left(\frac{rK}{L}\right)}{\left[\{Px(1-t)\}^{\frac{\alpha}{\alpha-1}} + (P_Y)^{\frac{\alpha}{\alpha-1}} - \left(\frac{W^A/\beta}{L}\right)^{\frac{\alpha}{\alpha-1}}\right]} \tag{11}$$

This is the aggregate supply function of child labour. We will now analyze its properties. To find out the impact of change of various policy parameters on the incidence of child labour supply we proceed as follows:

Now,

$$\log L_c^s = \log V - \log Z$$

where V and Z are respectively the numerator and denominator of Eq. (11).

Case 1: Impact of adult wage rate change on supply of child labour

$$\begin{aligned}\frac{1}{L_c^s}\frac{\partial L_c^s}{\partial W^A} = \frac{1}{V}\left[\frac{1}{(\beta L)^{\frac{1}{\alpha-1}}}\left(\frac{\alpha}{\alpha-1}\right)(W^A)^{\frac{1}{\alpha-1}} + \frac{1}{(\beta L)^{\frac{1}{\alpha-1}}}\left(\frac{rK}{L}\right)\left(\frac{1}{\alpha-1}\right)(W^A)^{\frac{2-\alpha}{\alpha-1}}\right] \\ + \frac{1}{Z}\left(\frac{\alpha}{\alpha-1}\right)(W^A)^{\frac{1}{\alpha-1}}\left(\frac{1}{(\beta L)}\right)^{\frac{\alpha}{\alpha-1}} < 0\end{aligned} \tag{12}$$

Higher is the adult wage rate, lower will be the tendency of adult workers to send their siblings to job market.

Case 2: Impact of rental rate change on supply of child labour

$$\frac{1}{L_C^S}\frac{\partial L_C^S}{\partial r} = \frac{1}{V}\left(\frac{W^A}{\beta L}\right)^{\frac{1}{\alpha-1}}\left(\frac{K}{L}\right) > 0 \tag{13}$$

Higher the rental rate, higher will be the utilization of adult labour in per unit *Y* production. Assuming ceteris paribus, it implies an increase in child labour supply in export sector. The supply of child labour has to go up in *X* sector to maintain the output of *X* sector unaffected.

Case 3: Impact of expansion of capital stock on child labour supply

$$\frac{1}{L_C^S}\frac{\partial L_C^S}{\partial K} = \frac{1}{V}\left(\frac{W^A}{\beta L}\right)^{\frac{1}{\alpha-1}}\left(\frac{r}{L}\right) > 0 \tag{14}$$

Higher is the stock of capital, higher will be the recruitment of adult labour in import-competing sector as we have assumed full employment of capital. Assuming ceteris paribus in export sector, it implies increase in child labour supply.

Case 4: Impact of expansion of adult labour endowment

$$\begin{aligned}\frac{1}{L_C^S}\frac{\partial L_C^S}{\partial L} = -\frac{1}{V}\left[\frac{(W^A)^{\left(\frac{\alpha}{\alpha-1}\right)}}{(\beta)^{\frac{1}{\alpha-1}}}\left(\frac{1}{\alpha-1}\right)L^{\left(\frac{-\alpha}{\alpha-1}\right)} + \left(\frac{W^A}{\beta}\right)^{\frac{1}{\alpha-1}}\left(\frac{1}{\alpha-1}\right)\frac{rK}{L}L^{\left(\frac{-\alpha}{\alpha-1}\right)}\left(\frac{W^A}{\beta L}\right)^{\frac{1}{\alpha-1}}\frac{rK}{L^2}\right] \\ -\frac{1}{Z}\left[\left(\frac{W^A}{\beta}\right)^{\frac{\alpha}{\alpha-1}}\left(\frac{\alpha}{\alpha-1}\right)(L)^{\frac{1-2\alpha}{\alpha-1}}\right] > 0\end{aligned} \tag{15}$$

As the adult labour endowment increases, supply of child labour in the economy increases.

Case 5: Impact of increase in tariff rate

$$\frac{1}{L_c^s}\frac{\partial L_c^s}{\partial t} = \frac{1}{Z}\left(\frac{\alpha}{\alpha-1}\right)\{Px(1-t)\}^{\frac{1}{\alpha-1}}(-1) < 0 \tag{16}$$

Higher is the rate of tariff imposed by the rest of the world on exported product of the small open economy, lower will be the incentive for the producers to produce the product. Hence, child labour supply will fall.

Thus the aggregate child labour supply function in functional form can be written as

$$L_C^S = L_C^S(W^A, r, K, L, t)\,\frac{\partial L_C^S}{\partial W^A} < 0, \frac{\partial L_C^S}{\partial r} > 0, \frac{\partial L_C^S}{\partial K} > 0, \frac{\partial L_C^S}{\partial L} > 0, \frac{\partial L_C^S}{\partial t} < 0 \quad (17)$$

4 The General Equilibrium Analysis

Given the assumptions of perfectly competitive markets, the following equations display the zero profit conditions for the economy:

$$a_{LaX} W^A + a_{LcX}(W^A/\beta) = P_X(1-t) \quad (18)$$

$$a_{LaY} W^A + a_{KY} r = P_Y \quad (19)$$

Child and adult labour are two different factors of production with different wage rates, W^A and W^A/β.

The full employment conditions for the economy are:

$$a_{LaX} X + a_{LaY} Y = L \quad (20)$$

$$a_{LcX} X = L_C^S = L_C^S(W^A, r, L, t) \quad (21)$$

$$a_{KY} Y = K \quad (22)$$

There are five endogenous variables: W^A, r, X, Y and L_C^S and five Eqs. (18–22). The parameters in the model are: P_X, P_Y, L, K, t.

We note that the system possesses decomposition property since the unknown input prices W^A (hence W^C) and r can be solved from the price system alone independent of the output system. Once the factor prices are known factor coefficients a_{ij}'s are also known. Y is solved from Eq. (22). Substituting the value of Y in (20) we shall solve for X. L_C^S will be solved from (21).

5 Comparative Statics

In this section, we shall discuss the effectiveness of trade restrictive policy on the incidence of child labour. Following Maskus (1997), the impact of stringent trade restriction in our model is captured by an increase in the tariff rate that the rest of the world imposes on the exported product of the small open economy.

5.1 Imposition of Stringent Trade Restriction by Rest of the World on the Exported Product of the Small Open Economy and Effect on Factor Price and Child Labour Supply

To find out the impact of restrictive trade policy on child labour supply, we differentiate Eq. (17) with respect to tariff rate t and the following expression is obtained

$$\frac{\mathrm{d}L_C^S}{\mathrm{d}t} = \frac{\partial L_C^S}{\partial W^A}\frac{\mathrm{d}W^A}{\mathrm{d}t} + \frac{\partial L_C^S}{\partial r}\frac{\mathrm{d}r}{\mathrm{d}t} + \frac{\partial L_C^S}{\partial K}\frac{\mathrm{d}K}{\mathrm{d}t} + \frac{\partial L_C^S}{\partial L}\frac{\mathrm{d}L}{\mathrm{d}t} + \frac{\partial L_C^S}{\partial t} \tag{23}$$

Now we shall first discuss the impact on factor prices.
Taking total differentiation of (18) we get,

$$a_{LaX}\mathrm{d}W^A + W^A\mathrm{d}a_{LaX} + a_{LcX}\mathrm{d}(W^A/\beta) + (W^A/\beta)\mathrm{d}a_{LcX} = -P_X\mathrm{d}t$$

or,

$$\theta_{LaX}\widehat{W}^A + \theta_{LaX}\hat{a}_{LaX} + \theta_{LcX}\widehat{W}^A + \theta_{LcX}\hat{a}_{LcX} = -\mathrm{d}t \tag{24}$$

Now the cost minimization condition of the producer entails that

$$\begin{aligned} &\mathrm{d}a_{LcX}/\mathrm{d}a_{LaX} = -\left(W^A/W^C\right) \\ &or, W^C\mathrm{d}a_{LcX} + W^A\mathrm{d}a_{LaX} = 0 \\ &or, \theta_{LcX}\hat{a}_{LcX} + \theta_{LaX}\hat{a}_{LaX} = 0 \end{aligned} \tag{25}$$

∴ From (24) we get, $\widehat{W}^A(\theta_{LcX} + \theta_{LaX}) = -\mathrm{d}t$
or,

$$\widehat{W}^A = -\mathrm{d}t(\text{since}\,\theta_{LcX} + \theta_{LaX} = 1) \tag{26}$$

Since,

$$W^C = (W^A/\beta) \tag{27}$$

$$\begin{aligned} &\therefore \mathrm{d}W^C = (1/\beta)\mathrm{d}W^A \\ &\therefore \widehat{W}^C = \widehat{W}^A = -\mathrm{d}t < 0 \end{aligned} \tag{28}$$

Differentiating (19) we get,

$$\begin{aligned} a_{LaY}\mathrm{d}W^A + W^A\mathrm{d}a_{LaY} + a_{KY}\mathrm{d}r + r\mathrm{d}a_{KY} = 0 \\ \mathrm{d}or, \theta_{LaY}\widehat{W}^A + \theta_{LaY}\hat{a}_{LaY} + \theta_{KY}\hat{r} + \theta_{KY}\hat{a}_{KY} = 0 \end{aligned} \tag{29}$$

From condition of cost minimization,

$$\theta_{LaY}\hat{a}_{LaY} + \theta_{KY}\hat{a}_{KY} = 0 \tag{30}$$

∴ From (29) we can write,

$$\theta_{LaY}\widehat{W}^A + \theta_{KY}\hat{r} = 0$$

Or

$$\hat{r} = (\theta_{LaY}/\theta_{KY})\mathrm{d}t > 0 \tag{31}$$

Proposition 1 *Due to the imposition of trade restriction on the exported product of the small open economy by the rest of the world both adult and child workers will be adversely affected and capital owners will gain.*

The intuitive explanation behind Proposition 1 is straightforward: Due to fall in effective producer's price in the exportable sector they will offer less to the factors of production engaged in the sector. It indicates a shock in labour (both adult and child) market. On the other hand, the price of import-competing goods remains unchanged. Hence to ensure normal profit producers must pay more to the capital owners.

Differentiating (17) we get,

$$\mathrm{d}L_C^S = \frac{\partial L_C^S}{\partial W^A}\mathrm{d}W^A + \frac{\partial L_C^S}{\partial r}\mathrm{d}r + \frac{\partial L_C^S}{\partial t}\mathrm{d}t \tag{32}$$

or,

$$\widehat{L}_C^S = \left(\frac{\partial L_C^S}{\partial W^A}\frac{W^A}{L_C^S}\right)\widehat{W}^A + \left(\frac{\partial L_C^S}{\partial r}\frac{r}{L_C^S}\right)\hat{r} + \left(\frac{\partial L_C^S}{\partial t}\frac{t}{L_C^S}\right)\hat{t} \tag{33}$$

$$= E_{W^A}\widehat{W}^A + Er\hat{r} + Et\hat{r} \gtrless 0 \tag{34}$$

Proposition 2 *If we assume adult worker's utility function to be additive separable in nature then imposition of trade restriction on child labour using exported product by rest of the world may or may not have perverse impact on incidence of child labour supply.*

5.2 Alternative Specification of Utility Function and Effect on Child Labour Supply

The result will be different if there is change in the form of utility function. Let us consider the following specific Cob–Douglas type of utility function:

$$U = AX^a Y^b / (L_C^S)^C \text{ where } \quad A > 0, \quad 0 < a, b, c < 1 \tag{35}$$

It satisfies all the standard properties and it is homogeneous of degree (a + b − c), where $(a + b - c) \leq 1$ or ≥ 1.

The adult worker maximizes the utility function subject to the budget constraint

$$Px(1-t)X + P_Y Y = \left[\left\{\left(\frac{W^A}{\beta}\right) L_c^s + W^A L + rK\right\} \Big/ L\right] \tag{36}$$

The Lagrangian expression for the constrained optimization problem is

$$M = \left[AX^a Y^b / \left(L_C^S\right)^C\right] + \lambda\left[P_x(1-t)X + P_Y Y - \left\{\left(\frac{W^A}{\beta}\right) L_c^s + W^A L + rK\right\} \Big/ L\right] \tag{36a}$$

The first-order conditions (FOCs) for maximization are

$$\delta M / \delta X = AaX^{a-1} Y^b (L_c^s)^{-c} + \lambda P_X (1-t) = 0 \tag{37}$$

$$\delta M / \delta Y = AbX^a Y^{b-1} (L_c^s)^{-c} + \lambda P_Y = 0 \tag{38}$$

$$\delta M / \delta L_c^s = -AcX^a Y^b (L_c^s)^{-c-1} - \lambda (W^A / \beta) L = 0 \tag{39}$$

$$\partial M / \partial \lambda = Px(1-t)X + P_Y Y - \left[\left\{\left(\frac{W^A}{\beta}\right) L_C^S + W^A L + rK\right\} \Big/ L\right] = 0 \tag{40}$$

From FOCs we get,

$$\left.\begin{aligned} X &= (a/c)\left(\frac{(W^A/\beta) L_c^s}{L}\right) / P_X (1-t) \\ Y &= (b/c)\left(\frac{(W^A/\beta) L_c^s}{L}\right) / P_Y \end{aligned}\right] \tag{41}$$

Substituting the values of X and Y in the budget constraint

$$P_X(1-t)(a/c)\left(\frac{(W^A/\beta)L_c^s}{L}\right)/P_X(1-t)$$

$$+P_Y(b/c)\left(\frac{(W^A/\beta)L_c^s}{L}\right)P_Y L_C^S/P_Y = \left(\frac{(W^A/\beta)L_c^s}{L}\right)+W^A+\frac{rK}{L} \tag{42}$$

$$\therefore L_c^s = \frac{\beta\left(1+\frac{rK}{LW^A}\right)}{\left(\frac{a}{c}+\frac{b}{c}-\frac{1}{L}\right)} = \frac{\beta\left(LW^A+rK\right)c}{W^A[(a+b)-c]}$$

This is the aggregate supply function of child labour. We shall now analyze its properties.

$$\log L_c^s = \log\beta + \log c + \log(LW^A+rK) - \log W^A - \log\{(a+b)-c\} \tag{43}$$

$$\therefore \frac{1}{L_c^s}\frac{\partial L_c^s}{\partial W^A} = \frac{L}{LW^A+rK} - \frac{1}{W^A} = \frac{-rK}{\left(LW^A+rK\right)} < 0 \tag{44}$$

$$\frac{1}{L_c^s}\frac{\partial L_c^s}{\partial r} = \frac{K}{LW^A+rK} > 0 \tag{45}$$

$$\frac{1}{L_c^s}\frac{\partial L_c^s}{\partial K} = \frac{r}{LW^A+rK} > 0 \tag{46}$$

$$\frac{1}{L_c^s}\frac{\partial L_c^s}{\partial L} = \frac{W^A}{LW^A+rK} > 0 \tag{47}$$

Hence the aggregate supply function in functional form can be written as

$$L_C^S = f(W^A, r, K, L), \text{where} \frac{\partial L_C^S}{\partial W^A} < 0, \frac{\partial L_C^S}{\partial r} > 0, \frac{\partial L_C^S}{\partial K} > 0, \frac{\partial L_C^S}{\partial L} > 0 \tag{48}$$

Now due to imposition of stringent trade restriction on the exported product of the small open economy by the rest of the world, there will be change in aggregate supply of child labour. To see how it changes, we differentiate (48).

$$\mathrm{d}L_C^S = \frac{\partial L_C^S}{\partial W^A}\mathrm{d}W^A + \frac{\partial L_C^S}{\partial r}\mathrm{d}r + \frac{\partial L_C^S}{\partial K}\mathrm{d}K + \frac{\partial L_C^S}{\partial L}\mathrm{d}L \tag{49}$$

In our model dK = dL = 0,

$$\therefore \mathrm{d}L_C^S \frac{\partial L_C^S}{\partial W^A}\mathrm{d}W^A + \frac{\partial L_C^S}{\partial r}\mathrm{d}r \tag{50}$$

or,

$$\widehat{L}_C^S = \left(\frac{\partial L_C^S}{\partial W^A}\frac{W^A}{L_C^S}\right)\widehat{W}^A + \left(\frac{\partial L_C^S}{\partial r}\frac{r}{L_C^S}\right)\hat{r} > 0 \tag{51}$$

Proposition 3 *Due to imposition of trade restriction on the exported product of the small open economy which uses child labour a paradoxical result is obtained—supply of child labour goes up unambiguously.*

6 Concluding Remarks

In this paper, we have considered a competitive two sector general equilibrium model of a small open less developed economy which exports products using child labour. We have examined the effectiveness of trade restrictive policies on the incidence of child labour . we have observed that (a) both adult and child worker will be adversely affected and capital owners will gain, but (b) incidence of child labour supply may or may not be perversely affected, if the adult worker's utility function is assumed to be additive separable in nature, and(c) may in fact increase the supply of child labour unambiguously if the adult worker's utility function is assumed to be Cob–Douglas in nature. This result thus challenges the popular view in favour of imposing trade sanctions on the import of those goods from the developed countries which are produced using child labour.

References

Baland JM, Robinson JA (2000) Is child labour inefficient? J Polit Econ 108(4):663–679
Basu K (1999) Child labour: cause, consequence, and cure, with remarks on international labour standards. J Econ Lit 37(3):1083–1119
Basu K (2000) The intriguing relationship between adult minimum wage and child labour. Econ J 110(462):C50–C61
Basu K, Van PH (1998) The economics of child labour. American Econ Rev 88(3):412–427
Bhagwati J (1995) Trade liberalization and 'Fair Trade' demands: addressing the environment and labour standards issues. World Econ 18:745–759
Dixit A (2000) Comment on A transaction cost politics analysis of International Child labour standards. In: A.V. Deardonff and R. Stern (eds) Social dimensions of U.S. Trade policies
Edmonds E, Pavcnik N (2002) Does globalization increase child labour? Evidence from vietnam. NBER working paper, No. 8760
Grootaert C, Kanbur R (1995) Child labour: an economic perspective. Int Labour Rev 134 (2):187–203
Gupta MR (2003) Child labour and economic development: a theoretical analysis. Arthaniti 2:97–107
ILO (1996) Economically active populations—estimates and projections, 1956–2010. ILO, Geneva

ILO (2014) IPEC action against child labour 2012-2013: Progress and future priorities
Jafarey S, Lahiri S (2002) Will trade sanctions reduce child labour? the role of credit markets. J Dev Econ 68(1):137–156
Jones RW (1971) A three-factor model in theory, trade, and history. In: Jagdish B et al (eds) Trade, balance of payments and growth: essays in honour of Charles P. Kindleberger
Maskus KE (1997) Should core labour standards be imposed through international trade policy? World bank, policy research working paper, No. 1817
Ranjan P (2001) Credit constraints and the phenomenon of child labour. J Dev Econ 64(1):81–102
Ravi A (2001) Combating child labour with labels: case of rugmark. Econ Polit Wkly 1141–1147
Ray R (2002) The determinants of child labour and child schooling in ghana. J Afr Econ 11(4): 561–590
Stein E, Davis J (eds) (1940) Labour problem in America. F. Forraer and Rinchar

Part IV
Issues Related to Foreign Investment Flows

The Determinants of Foreign Direct Investment: An Analytical Survey

Chaitali Sinha and Kunal Sen

Abstract Investments undertaken by multinational corporations (MNCs) can be regional or global. In the recent years, there has been spectacular growth in the flow of global capital in developing economies accompanied by a significant increase in capital outflows from the Global South. Though outward investment by emerging economies is not a new phenomenon, the past couple of decades have witnessed a surge in the quantity and also qualitative transformation in the pattern of their investment. The most important region among developing countries for foreign direct investment outflows is Asia, though there has been an increase in other developing regions as well. In this chapter, we present an analytical survey of the literature around the various issues that determine the location decision of global capital and also investigate the current changes in trends in these investment flows. We also explain the motivating factors behind the location decisions of southern multinationals, an area relatively new in the literature on foreign direct investment.

1 Introduction

International business activity in the form of multinational corporations (henceforth, MNCs) is not a recent phenomenon. These economic activities that have their roots in the nineteenth century included foreign direct investment (henceforth, FDI), joint ventures and strategic alliances, among other forms of internationalization. There is a long-standing notion among the policy makers that FDI is much more inducive to long run economic growth and development when compared to other forms of foreign capital flows. However, despite the presence of FDI, most of the foreign investments till late 1940s were in the form of portfolio investment. The volume of

C. Sinha · K. Sen (✉)
Institute for Development Policy and Management (IDPM), School of Environment, Education and Development, University of Manchester, Manchester, UK
e-mail: kunal.sen@manchester.ac.uk

C. Sinha
e-mail: chai_sinha@yahoo.co.in

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_16

FDI grew substantially after the World War II and concentrated its activities towards knowledge-based production rather than in primary goods. In the early 1960s, the United States accounted for about three-fifth of the total FDI of the market economies and was followed by the United Kingdom and other OECD countries. However, changing patterns of industrial production was reflected in the declining shares of US and UK in FDI stocks and rise of Germany and Japan between 1967 and 1976. Big business houses shifted their interest from traditional locations such as Canada, Latin America, Ex-colonial territories to the newly industrialized areas such as South-East Asia along with a shift in the nature of international production. Rather than extracting natural resources overseas, MNCs started concentrating on production specialization (both horizontal and vertical) to take advantage of difference in endowment across nations, scale economies and integrated markets.

Past couple of decades have experienced spectacular growth in the flow of global capital into developing countries, including many of the debt-stricken Latin American countries. Figures 1 and 2 show per cent share of FDI inflows for different regions since 1970. With global capital flows growing at a faster rate than the

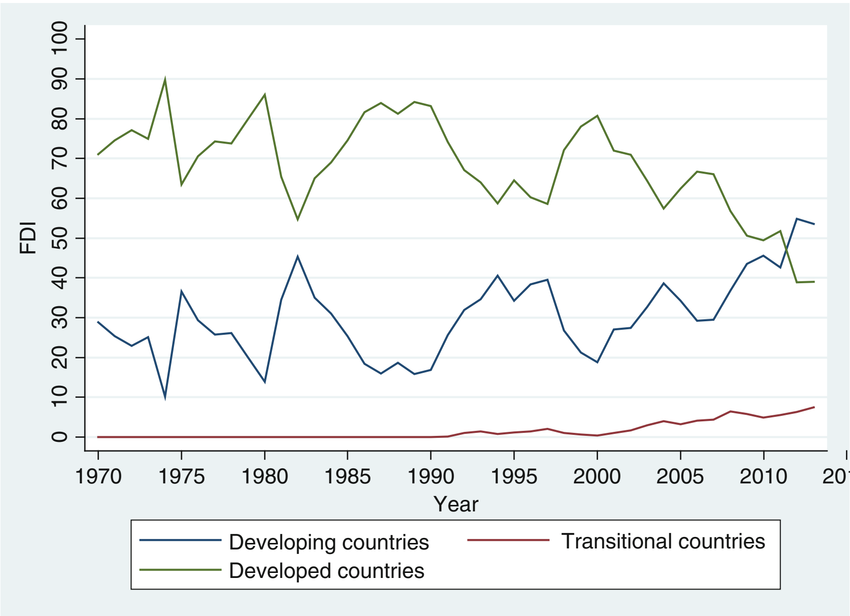


Fig. 1 Per cent share of world FDI inflows for developing, developed and transitional countries. *Source* UNCTAD and authors' calculation

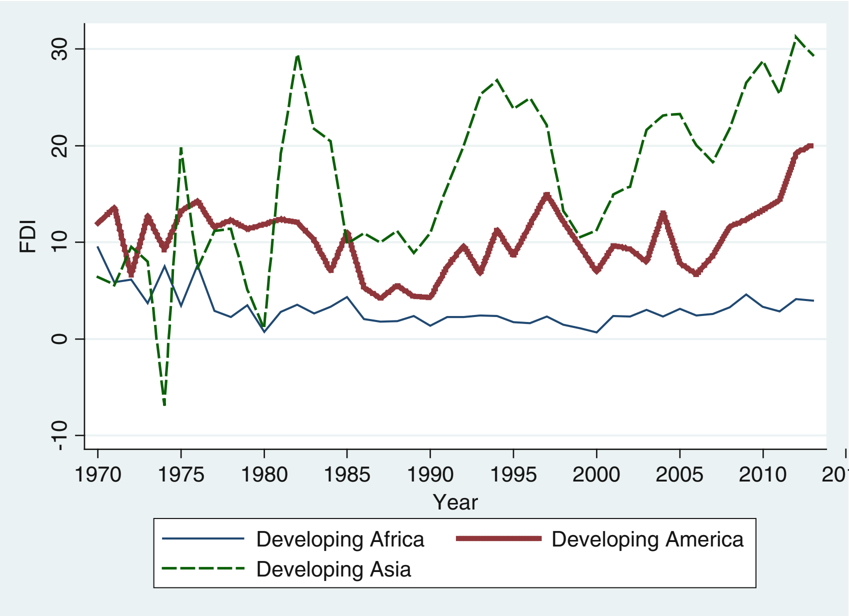


Fig. 2 Per cent share of world FDI inflows for developing Africa, developing Asia and developing America. *Source* UNCTAD and authors' calculation

global trade, it remains an open question that what pulls FDI into the emerging economies that are often protected.

It is not difficult to understand that FDI will flow to the countries with relatively stable economic, political and social conditions accompanied with strong institutions. However, there is very little evidence to support this view.[1] Structurally weak economies like least developed countries (LDCs), landlocked developing countries (LLDCs) and small island developing states (SIDS) experienced increase in FDI inflows by 29, 54 and 32 %, respectively, in 2008 (Fig. 3). FDI outflows from developing countries have also experienced major increase. This surge in outward investment is mainly due to cross-border mergers and acquisitions (for example, Cemex, from Mexico has become the largest cement producer in US by acquisition; Italian company Wind was purchased by Egyptian EMNC Orascom, etc.). Flows from

[1]Though, internal factors of FDI receiving country have received huge attention in the present literature (see Blonigen 2005), there exists a large body of work that examines the importance of external forces in driving foreign capital, mainly debt and portfolio flows to the rising economies (see Calvo et al. 1993; Fernandez-Arias 1996; Reinhart and Montiel 2001).

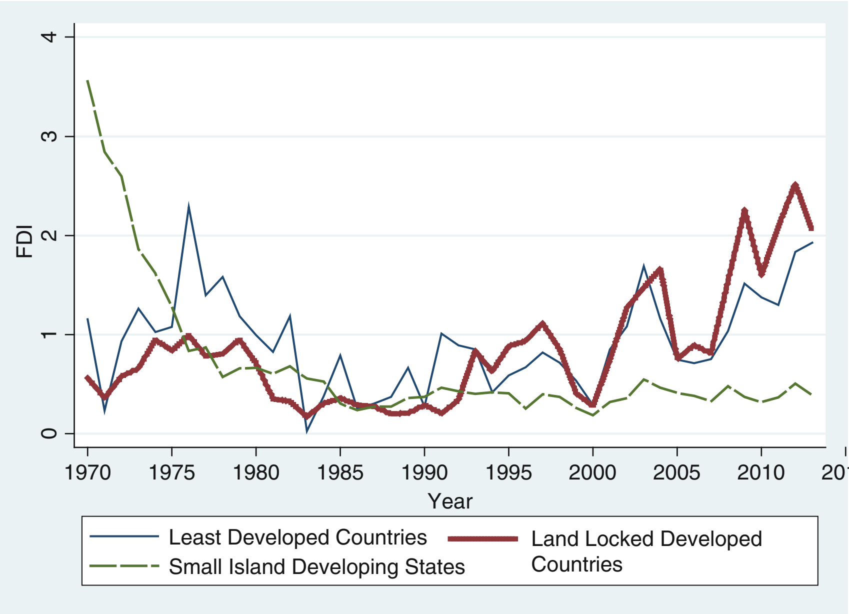


Fig. 3 Per cent share of world FDI inflows to LDCs, LLDC and SIDS. *Source* UNCTAD and authors' calculation

the developing countries have increased from 2.7 % of the total global outflows in 1989–1991 to 13 % of the total global outflows in 2007.[2]

However, flow of private capital has been both cyclical and inconsistent. Policies such as financial openness, adopted by many of the developing economies have significantly contributed to the surge in the inflow of foreign capital and have given greater exposure to the global financial shock. This inflow of private capital has been a cause of global liquidity, accompanied by growing commodity prices, declining interest rate, better economic fundamentals and market-oriented reforms adopted by many of the rising economies. However, economists are yet to reach a firm conclusion whether these capital inflows to the developing economies has been

[2]Investment by multinational corporations can be regional or global. Data on MNCs from emerging markets show that they are likely to invest in their own region or in other developing countries with whom they are familiar through trade, or have ethnic and cultural ties (for example initially Russia invested in other countries of former Soviet Union, India and China mainly invested in other Asian countries, South Africa's investment abroad was mostly to other countries of Southern Africa and Chile, Brazil and Argentina invested in other countries of their region). This surge in intraregional or South-South FDI flows since early 2000s was mainly due to availability of Petrobras in Argentina, Bolivia and Venezuela, giving access to oil and gas reserves; state policy of regional energy integration in Argentina, Brazil, and Cuba; and retreating some of the global MNCs from Latin America during the early 2000s that gave local firms the opportunity to increase their activities in the region.

a result of the deteriorating macroeconomic conditions of the developed world known as *pushed* factors or have been *pulled* by improving domestic conditions of the developing economies. There are many studies that come across the relationship between FDI and several macroeconomic variables such as market size of the host countries, economic and political stability, government policies, infrastructure, degree of openness of the host country, quality of institutions, absorptive capacity of the local firms, human capital, cost of labour, etc. However, increase in foreign assets and liabilities of many of the developing countries were results of improved current account balance that reduced their foreign debt and helped building international reserves. With the decline in the world interest rate, the debt servicing burden of the developing countries declined substantially. This unprecedented rise in the holding of foreign reserves by the developing countries (China showed a huge accumulation of reserve), commonly known as *self-insurance*, helped them to fight back the crisis of the late 1990s. However, this unparalleled rise in the flow of private capital to the developing countries that reached its peak during 2007–2008 came to a *sudden stop* or even reversed its direction and flew back to the developed countries where the epicentres of the global financial crisis existed. The world economy started recovering since spring 2009 with the help of the support laid by the central banks of the developed countries. However, developing countries saw another surge in the inflow of global capital after mid-2009 followed by another reversal of flows as an aftermath of worsening European crisis of 2011. Regional analysis of data shows, FDI flows to developing countries reached to $778 billion, a share of 54 % of the total global inflows (UNCTAD 2014). Major developing areas such as developing Asia, Africa, Latin America and Caribbean experienced a major rise in the share of global flows. Africa experienced a growth of 4 % in inflow of foreign capital mostly due to intra-African flows. Developing Asia saw a rise of 3 % and the Latin America and the Caribbean experienced an overall positive growth.

With the increasing trend of foreign capital pouring in developing economies and outflows from developing economies rising, the obvious question arises is how the *push* and the *pull* factors operate to determine the location choice of FDI.

This chapter surveys the literature around the various factors that determine flow of FDI, investigates the motives and strategies of MNCs that determine the location decision of FDI and also investigates the current surge of outflows from the Global South.

2 Changing Map of FDI in the Recent Years

With the growing integration of the global capital markets, FDI grew significantly during the 1990s at a rate faster than global economic growth and trade. Data shows that world FDI flows that increased by an average of 13 % a year during 1990–1997, saw an average increase of 50 % during 1998–2000 due to mergers and acquisitions. Global inward FDI flows that rose from US$54.1 billion in 1980, reached US$207.7 billion in 1990 and US$1,401.5 billion in 2000 and felled to US $0.7 trillion in 2001 as a result of sharp decline in mergers and acquisition. By 2003 it had fallen to US$565.7 billion before rising again to US$2100 billion in 2007. According to UNCTAD (2014), developing and transitional countries together invested 39 % of total global FDI outflows in 2013. Initially, Argentina, Brazil, Hong Kong, India, Korea, Singapore and Taiwan were the major sources of emerging country's FDI. However, since late 1980s these countries were joined by Chile, China, Egypt, Malaysia, Mexico, Russia, South Africa, Thailand and Turkey.

Three major economic groups, the developed, the developing and the transition economies of South-East Europe and Commonwealth of Independent States (CIS) were affected by the Global financial crisis of 2008 differently and this was reflected by the respective falls in their FDI inflows. The developed countries saw a 29 % decline in inward FDI flows in 2008 which was mostly due to drop in cross-border mergers and acquisition (M&A) sales. Most of them suffered a downfall as host economy due to the global crisis, except US. The developing countries survived the crisis of 2008 as they were not tightly interlinked with the banking system of US and Europe that were badly hit by the disaster. With their economic growth remaining robust and commodity prices rising, the developing countries continued to face growth in inward FDI, however, at a rate slower than the previous year. While the manufacturing and the service sectors were the worst hit, the primary sector saw a rise in FDI mainly due to participation of large companies from developing countries, especially from China. For the least developed countries, Africa received US$88 billion in 2008 amid the global economic and financial crisis and most of these FDIs that were directed from developed countries were concentrated in natural resource-based industries (Fig. 4). Many of the African countries adopted policy measures to make environment favourable to FDI, however, the recent picture across different African regions remain mixed. On the other hand, though South, East and South-East Asia collectively experienced a huge growth in FDI inflow of 17 %, reaching a new high, in 2008, the picture varied significantly among different regions and for West Asia the picture was rather mixed[3] (Fig. 5).

[3]Whereas, inflows slightly dropped in Malaysia and Thailand, it declined sharply in Singapore and Taiwan province of China. However, China, India, Republic of Korea and Hong Kong (China) experienced an increase in FDI inflow. Though in total, West Asia experienced a significant increase of 16 % in FDI inflows, this was mainly due to major growth experienced by real estate, petrochemicals and oil refinery industries of Saudi Arabia, as the major players like Turkey and United Arab Emirates received major set back. A similar picture was found in Latin America and the Caribbean where there was a 13 % increase in total FDI flows in 2008. With an unequal distribution of the inflows in different regions, natural resource-based industries were the main recipients, whereas manufacturing sector observed a drop due to sharp decline in the flows to Central America and the Caribbean. South-East Europe and Commonwealth of Island (CIS) received US$114 billion in the year, with Russian Federation, Kazakhstan and Ukraine being the major players received nearly 84 % of the total inflow. This vast area received a record level of FDI inflows in spite of regional conflicts in some of its regions.

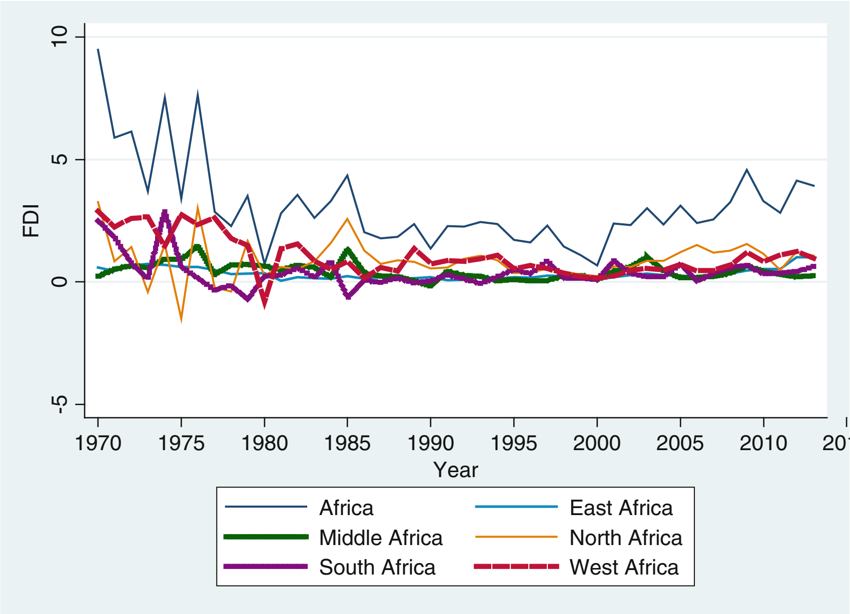


Fig. 4 Per cent share of world FDI inflows for different parts of developing Africa. *Source* UNCTAD and authors' calculation

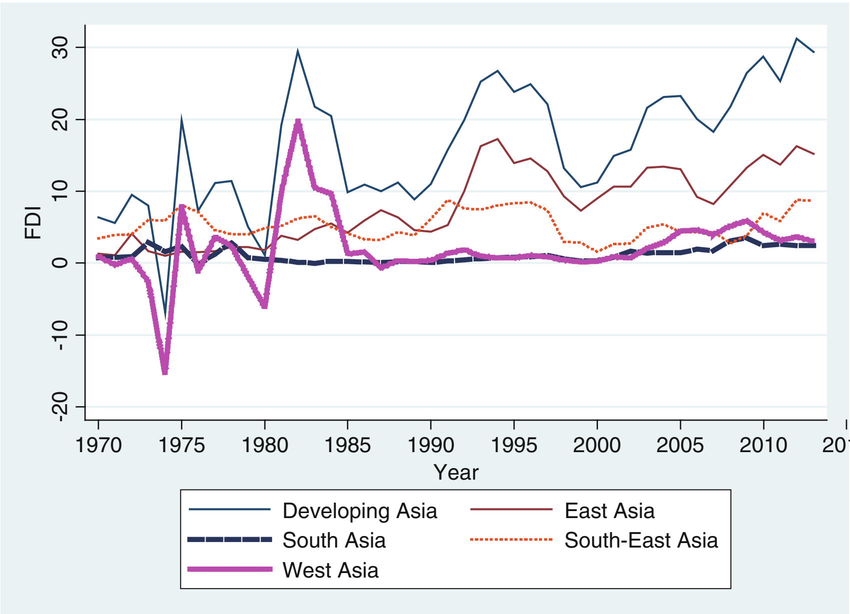


Fig. 5 Per cent share of world FDI inflows for different parts of developing Asia. *Source* UNCTAD and authors' calculation

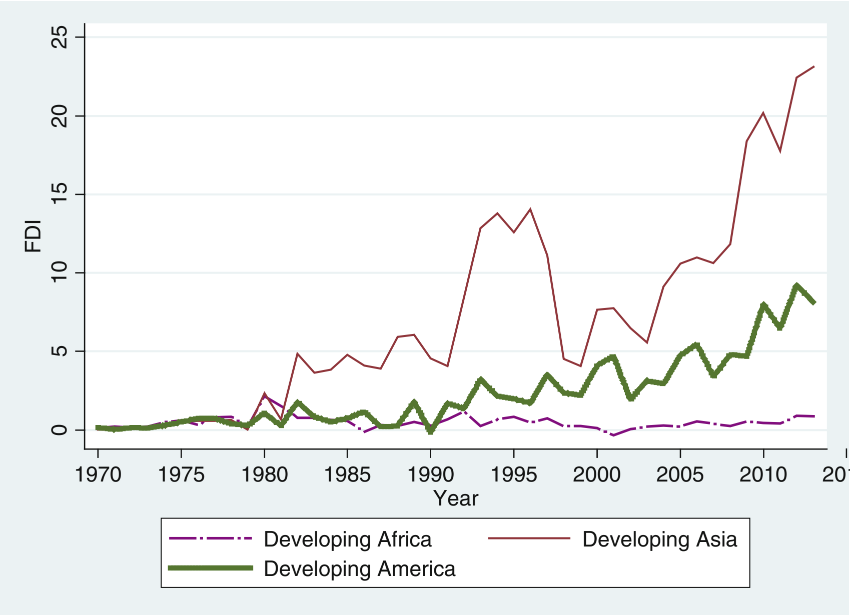


Fig. 6 Per cent share of world FDI outflow for developing Asia, developing America and developing Africa. *Source* UNCTAD and authors' calculation

As the world's major economies were badly affected by the increasing downturn in economic and financial crisis resulting in falling profits and declining reinvestments and rechannelling of loans from foreign affiliates to the headquarters of Transnational Corporations (henceforth, TNCs), FDI flows from developed countries fell by 17 % in 2008. These dramatic changes that occurred in 2008 changed the relative ranking of the host and the source countries in the world. While US maintained its top position both as a host and a source country, United Kingdom lost its position as the top host and source country in Europe. Japan improved its position in outward investment and many developing and transitional countries emerged as large recipients and investors of FDI in 2008. Figure 6 shows outflow of capital from developing Asia, Africa and America.

Outflows from West Asia declined by 30 % in 2008 mainly due to significant fall in the value of cross-border M&A purchases by West Asian TNCs. In contrast, FDI outflows from Latin America and the Caribbean increased by 22 %. This was mainly due to rising flow of FDI from South America that counterbalanced the decline in the outflows from Central America and the Caribbean. TNCs of Russian Federation continued to maintain their lead position. In addition, FDI flow from South, East and South-East Asia increased by 7 %, mainly due to huge outflow of FDI from China, though many of the countries of this region slowed down during 2009.

Most of these flows from the major economies of this region were due to relatively high economic growth and growing foreign reserves originated from trade surpluses and sovereign wealth funds (SWFs). Moreover, growing competition amongst the domestic firms; saturated or limited markets and improved institutional support contributed to the growth of FDI from this region. Most of the outward investment from this region is intraregional (for example in 2007, 40 % foreign investment by Temasek from Singapore were in Asia; Khazanah Malaysia, a Malaysian SWF, invested significantly in Malaysian companies such as UEM, Telecom Malaysia International, Opus Group Berhad and Bumiputra Commerce Bank). However, recently a growing number of developed countries are also receiving FDI from this region as a part of efforts of the Asian firms to get hold of strategic assets abroad. TNCs from East Asia are acquiring firms of the developed countries mainly which are based in United States because of weak dollar and lower asset prices of these companies.

Chinese overseas investment, mostly in extractive industries, has particularly focused on acquiring strategic assets outside Asia, mainly in developed countries, Africa and Latin America. Another major player, India, has been investing in both the developing and the developed countries, particularly in pharmaceuticals, extractive industries, information technology and other business services. Investment by Singapore firm Temasek Holdings in Merrill Lynch (United States), acquisition of Jaguar Cars Ltd. (United Kingdom) by Tata Motors Ltd. (India), overseas acquisition of Anglo-Dutch firm Corus by Tata group (India) are some of the largest deals carried out in the recent years. Outward investment by different Asian economies is shown in Fig. 7.

The year 2009 saw the Russian Federation to be the largest source of outward investor of FDI from the whole region. With rising number of Mexican and Brazilian companies expanding mainly in developed countries, the outward flow of FDI from Latin America and Caribbean increased to US$48 billion in 2003–2009 annually. Per cent share of outward investment by some of the Latin American countries and Russian Federation and Kazakhstan in global FDI outflows is given in Fig. 8. Moreover, though outward investment from Africa as a whole suffered, investment from South Africa and North Africa continued to grow. Despite a gloomy picture throughout Asia, China maintained its outward investment mainly in non-financial sectors.

2.1 Recovery of Global Investment and New Trends in Outflows from the Global South

The first half of 2010 saw a modest, however, uneven recovery of global FDI from 2009 crisis. Amidst the increasing risk and uncertainties of post-crisis world, featured by the possibilities of sovereign debt crisis, rising inflation, fiscal and financial imbalance of many of the developed countries and overheating in emerging market

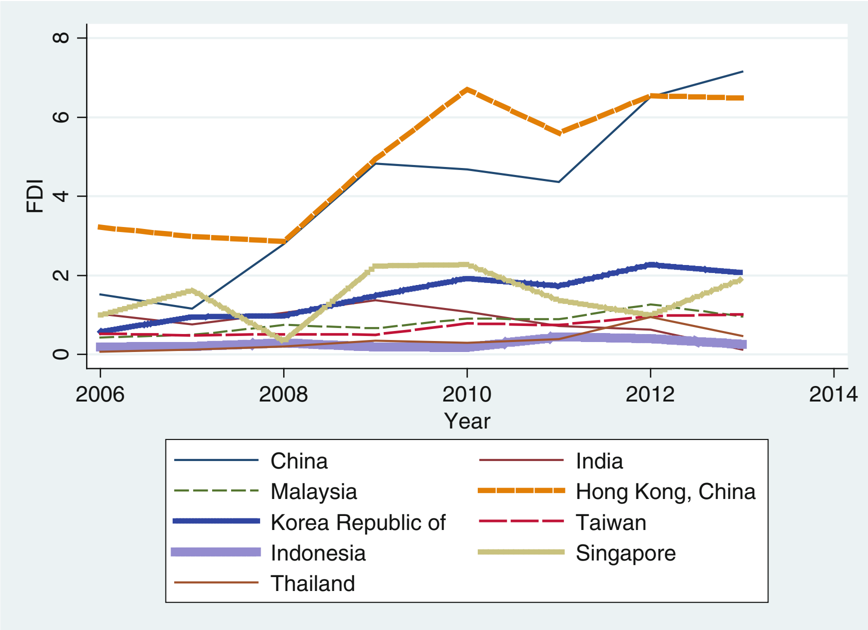


Fig. 7 Per cent share of different Asian economies in world FDI outflows. *Source* UNCTAD and authors' calculation

economies, though industrial output and world trade reached their pre-crisis level, FDI flows in 2010 remained at 15 and 37 % below their 2008 and 2007 peaks, respectively. The post-crisis period of 2010 saw developing countries to maintain their lead both as global recipients and investors of FDI. FDI in services (business services, finance transport and communication services) continued its downfall though at different paces.[4] 2010 also saw rise in FDI outflow by six developing and transitional countries that were among the top 20 investors.

Outflows from South-East Asia and West Asia saw a significant increase in 2011. While outflows from China and Hong Kong dropped; Singapore, Thailand and Indonesia saw a rise. Flows from India increased mainly due to increase in investment in overseas green field projects particularly in extractive industries;

[4]While FDI flows in the financial service sector declined the most, manufacturing industry backed most of the FDI investment. However, business-cycle-sensitive industries, for example metal and electronics sector suffered. Though chemical industry (including pharmaceutical industry) weathered away the crisis, others such as food, beverages, textile, tobacco and automobiles recovered in 2010. However, FDI in extractive services that were not affected by the crisis suffered a downfall in 2010.

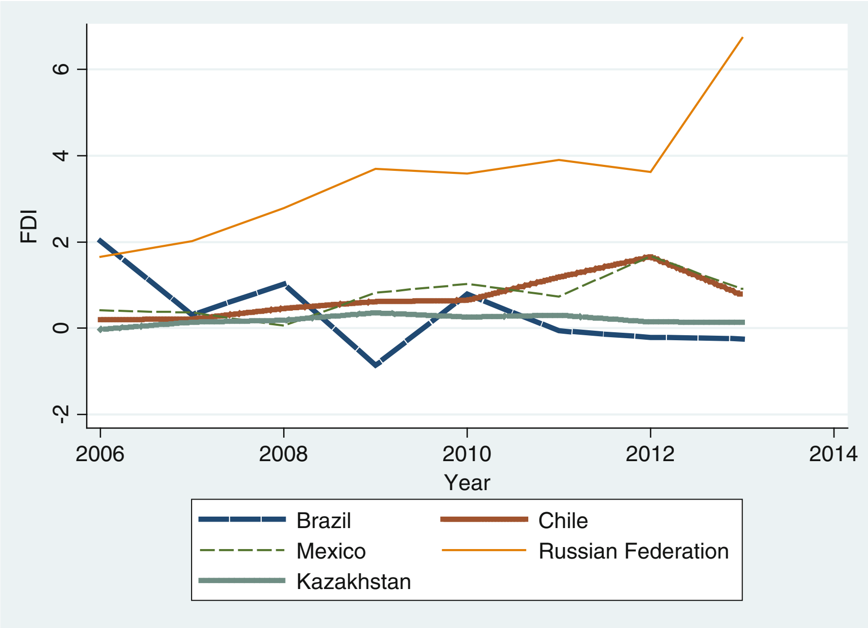


Fig. 8 Per cent share of world FDI outflows from Latin American and the Caribbean, Russian Federation and Kazakhstan. *Source* UNCTAD and authors' calculation

metal products and business services.[5] Political instability of Egypt and Libya and cancellation of big construction projects in the midst of global crisis in the Gulf Cooperation Council (GCC) lead to a fall in inflow of investment in those respective areas in 2011. In addition, political and social unrest contributed significantly to the fall in global investment in different parts of West Asia. However, Central America, South America, the Caribbean and South Asia saw high FDI inflows during 2011. In addition, improving political relationship between India and Pakistan generated attractive investment climate in the region of South Asia.

Global FDI outflows in 2012 increased by 5 % mainly due to investment from developing regions (whereas outflows from developing Asia and Africa increased, it declined for Latin America and the Caribbean) and transitional economies who continued to increase their outlay with the increasing trend in economic liberalization, growth and growing commodity prices. China and Hong Kong strengthened their positions as the two important sources of global FDI investors in 2013. Other parts of developing Asia saw mixed trends in outward flown in 2013. While

[5]FDI flows in 2011 rebounded in all the three sectors (primary, secondary and tertiary), with slightly higher growth rate for the primary and the service sectors compared to the manufacturing industry. The industries that contributed mainly to the rise in FDI in 2011 were mining, quarrying, petroleum; electricity, gas and water and transportation and communication services.

investment by TNCs from Republic of Korea declined by 5 %, for Taiwan it increased by 9 %. FDI flows from transitional countries increased significantly mainly due to TNCs from Russian Federation, followed by Kazakhstan and Azerbaijan. Drop in FDI outflows from developed countries in 2012 (Europe and North America experienced large downfalls in their outflows and Japan who was still able to maintain its position grew by 14 %) was marginally improved in 2013. Whereas outflows from Europe (doubling of reinvested earnings abroad and increase in intracompany loans helped Switzerland to become the largest outward investor in Europe in 2013 and outward flows from Italy, Netherlands and Spain rebounded in 2013) and Japan continued to grow in 2013, North America saw a 10 % decline due to negative outflow of intracompany loans from US. France, Germany and United Kingdom also saw a significant decline in outward flows in 2013. Inflows increased in North America by 23 % in 2013 mainly due to acquisitions by Asian companies (e.g. Canadian upstream oil and gas company, Nexen, was taken over by CNOOC of China; Sprint Nextel of US was taken over by Japanese telecommunication group Softbank and Smithfield of US was taken over by Chinese Shuanghui).

3 The Theoretical Background

According to neoclassical trade theory location decision of multinational firms has been guided by the theory of *comparative advantage* of the host country highlighted under the Ricardian or Heckscher–Ohlin–Vanek theoretical framework. According to the theory of *market imperfection* (Kindlerberger 1969; Eitemann et al. 2007), MNCs locate their production activities in other countries to take advantage of market imperfection by exploiting economies of scale, ownership advantage and government incentives. Moreover, by taking the advantage of market imperfection, MNCs safeguard their intangible assets (Buckley and Casson 1976; Hennart 1982; Shapiro 2006). In addition, according to Vernon (1966), MNCs' exporting of goods or setting up of production house in the foreign market is influenced by product life-cycle theory. However, the literature on the activities of MNCs, which was mainly based on the observations of international firms from triad (i.e. US, EU and Japan) is best explained by the most influential study called *eclectic paradigm* that was first proposed by Dunning (1981). Accordingly, the decision of the firms to expand overseas depends on three advantages: *Ownership advantage* (representing firm's specific resources to be exploited externally), *Location advantage* (representing host country's characteristics say endowment of natural resources) and *Internalization* (representing the opportunity to internalize firm specific advantages). The *eclectic paradigm* theory is the most comprehensive theory to provide rationalization of MNCs behaviour to invest in foreign countries. However, the conventional Ownership–Location–Internalization (OLI) paradigm of Dunning has been criticized on the ground that:

I. It could not explain internalization of the MNCs from the developing countries as these new firms do not possess the ownership advantage of superior technology or brand name that they can exploit in the foreign market. According to Mathews (2006, p. 17), "*Rather their international expansion has been undertaken as much for the search for new resources to underpin new strategic options, as it has been to exploit existing resources. This is why they have to expand quickly, to consolidate gains that are fleetingly won. This is why they tend to rely on partnerships and joint ventures, to reduce the high level of risk involved in their leveraged strategies*". The *challenger firms* or the *newcomer firms* from the developing countries take a very different perspective compared to the *incumbent firm*. According to (Aulakh 2007, p. 237), the *newcomer firms* from developing countries get motivated by "*…learning objectives that allow these firms to overcome the initial resource hurdles arising due to technological gaps and late mover disadvantages in international markets*". They acquire strategic assets like technologies and brands and raw materials that they lack by setting up linkages with the source firm overseas.
II. The OLI framework has been further criticized on the ground that it is a static paradigm and fails to explain the dynamism that captures the advancement of a firm's capabilities throughout time. The dynamic capability approach developed by Teece et al. (1997) is an extension of the resource-based view and is concerned with the knowledge-generating process at the firm level. Mathews (2002a, b) observed the dynamic behaviour of the *Dragon Multinationals* from the Asia Pacific region, and in a number of his successive works he has highlighted that the *resource-based approach* best fitted for this *newcomer firms* remains to be unexplained for quite a long time by the eclectic paradigm theory of Dunning.

This unexplained behaviour of the *newcomer/latecomer firms* were then best handled by an alternative framework called the *Linkage–Leverage–Learning* (LLL) developed by Mathews (2002a). Accordingly, *linkages*, formed by joint ventures or any other kind of collaboration with the *incumbent* or the foreign firms help the *latecomer firms* to access resources that they lack internally. After getting linked up these firms *leverage* their low production cost advantages and *learn* new foundation of competitive advantages and ways to operate them internationally.

However, the *LLL* approach has been criticized by Dunning (2006) and Narula (2006). Narula (2006) highlighted the fact that *LLL* approach focus mainly on the fast-growing economies of Asia Pacific region. Additionally, according to Dunning (2006), it has been found empirically that many of the *newcomer firms* possess some kind of competitive advantage with them that explains their process of internalization. However, Aykut and Goldstein (2006) and Mathews (2002b) have given another dimension to this view of Dunning by stating that enjoying some kind of competitive advantage by the *newcomer* firms, for example early awareness to plan their activities keeping in mind the global competition and partnership, would put them in the position to rapidly connect with the global world and *leverage* their resources for joint collaboration.

4 Empirical Findings

This section surveys the empirical literature around different *push* and *pull* factors behind FDI inflows; business cycles of the advanced countries as determinants of FDI inflows and also tries to identify the causes associated with emerging outflows from the Global South.

4.1 Domestic Pull *Factors as Determinant of FDI Inflows*

The vast empirical literature that focuses on the *pull* factors as the determinants of FDI use cross-country regressions to identify country-specific characteristics regarding market size, labour cost, political and institutional stability, government policies, etc., to attract FDI.

In this section, we discuss the literature around these factors that exhibit steady relationship with FDI in determining its location.

i. Market Size

Attractiveness of the host country's market, generally proxied by the country's GDP or per capita GDP, has been widely accepted in the empirical literature as a significant determinant of FDI flows (see Wafure and Nurudeen 2010; Artige and Nicolini 2006; Masayuki and Razafimahefa 2005; Jordaan 2004; Nonnenberg and Cardoso de Mendonça 2004; Chakrabarti 2001; Resmini 2000; Tsai 1994; Culem 1988; Schneider and Frey 1985; Schmitz and Bieri 1972; Bandera and White 1968 etc.). According to this hypothesis, larger size of the host country's market is associated with higher inflows of FDI as large markets are necessary for exploiting natural resources and economies of scale. Culem (1988) supported the market size hypothesis in a study of six industrialized countries, over the period 1969–1982. Similar results were reported by Papanastassiou and Pearce (1990), Swedenborg (1979) and Dunning (1980). Resmini (2000) found that for manufacturing FDI, countries of Central and Eastern Europe with huge population attract more FDI. Similar result was found by Bevan and Estrin (2000) where larger economies of the transitional countries attract more FDI. However, the relationship between market size and FDI is not unanimous. Whereas, Asiedu (2002); Jaspersen et al. (2000); Edwards (1990) found a negative relationship between real GDP per capita and FDI/GDP; Pistoresi (2000), Resmini (2000), Billington (1999), Shamsuddin (1994), Tsai (1994) and Schneider and Frey (1985) have found a positive relationship between market size and inward FDI.

However, there lie some conceptual problems with the market size variable (Chakrabarti 2001). Though most of the studies have used per capita real GDP as a measure of attractiveness of the host country's market and have shown significant positive relationship between market size and inward FDI; there exist some studies that have used absolute value of GDP as an alternative measure. Absolute value of

GDP of a developing country is a very poor measure of market attractiveness for the foreign investors as it really shows the size of the population rather than income. Moreover, use of GNP or GNP per capita as a measure of market size is also inappropriate in the context of foreign investment.

ii. Infrastructure Development

Another variable that attracts multinationals is good quality of infrastructure. Usually number of telephones per 1,000 population is taken as a measure of infrastructure development. Asiedu (2002); Loree and Guisinger (1995); Wheeler and Mody (1992) found that good quality infrastructure positively influences inflow of FDI.

iii. Wage or Labour Cost

One more important factor that determines FDI inflows, in spite of its controversial nature, is wages or labour cost (Chakrabarti 2001). Though, theoretically, cheap labour may be one of the most important causes of the multinationals venturing in the labour-abundant developing countries, no unanimity has been reached by the studies in exploring the role of wage in attracting FDI. Whereas, Wheeler and Mody (1992), Nankani (1979), Swedenborg (1979) found strong positive relationship between cheap labour and inflow of FDI; Pistoresi (2000), Shamsuddin (1994), Culem (1988), Schneider and Frey (1985), Flamm (1984), Saunders (1982), Goldsbrough (1979) found higher cost of labour to discourage inward FDI. Narula and Wakelin (2001) use unit labour cost in the manufacturing industry to show that cheap labour acts as an incentive for the US firms to invest in foreign countries. Lipsey (1999), Tsai (1994), Sader (1993), Lucas (1990); Gupta (1983) and Owen (1982) found statistically insignificant relationship between labour cost and inward FDI. Tsai (1994) got strong positive relationship between cheap labour and FDI for the period 1983–1986, but a very week association during the period 1975–1978.

iv. Openness

Openness, generally measured by the ratio of trade to GDP, is an important factor affecting the location choice of FDI. According to the openness hypothesis, a country's openness in international trade is a determining factor for FDI flows and the standard hypothesis is that openness encourages FDI. However, the impact depends on the type of foreign investment. If the investment is of market-seeking type then restricted trade can have positive influence on inward FDI flows. On the contrary, export orientation investment on the part of the multinational corporations may encourage them to locate in a more open economy. Leitão (2010), Hailu (2010), Quazi (2007), Sekkat and Veganzones-Varoudakis (2007), Nonnenberg and Cardoso de Mendonça (2004), Anyanwu and Erhijakpor (2004), Asiedu (2002), Noorbakhsh et al. (2001), Hausman and Fernandez-Arias (2000), Pistoresi (2000), Morisset (2000), Gastanaga et al. (1998), Edwards (1990), Culem (1988), Kravis and Lipsey (1982) found significant positive relationship between trade openness and inward FDI flows. Singh and Jun (1995) found openness to be very important in attracting FDI and also focused on the complementary relationship between the two. However, there are studies that show negative or week association between the variables. For example, Wheeler and Mody (1992) and Schmitz and Bieri (1972) found week link between trade openness and FDI. Yih Yun et al. (2000) reported negative relationship between the two variables.

v. Exchange Rate

According to the exchange rate hypothesis, weak currency discourages foreign investors to invest in that location. However, the findings in this area widely vary from being significantly positive to negative and insignificant. For example, Blonigen and Feenstra (1996), Blonigen (1995), Froot and Stein (1991) found strong negative relation between a country's exchange rate and FDI inflows; Tuman and Emmert (1999), Sader (1993) reported insignificant relationship between the two variables and Edwards (1990) found significant positive relationship.

vi. Taxes

Regarding the role of taxes in attracting FDI, there is no unanimity in the literature. Whereas, Swenson (1994) found positive significant relationship between the two variables; Billington (1999), Barrel and Pain (1998), Kemsley (1998), Cassou (1997), Loree and Guisinger (1995), Guisinger (1985), Hines and Rice (1994), Grubert and Mutti (1991), Hartman (1984) found significant negative association between corporate tax rate of the host country and inward FDI. Moreover, there are papers, for example Porcano and Price (1996), Jackson and Markowski (1995), Yulin and Reed (1995), Wheeler and Mody (1992) that found no significant relationship between the two variables.

vii. Human Capital

There is mixed evidence regarding the role played by human capital in an economy in attracting FDI. The relationship between MNCs and human capital is not straightforward. The relation follows two main directions. On the one hand, MNCs enter only when the critical level of human capital is high in the host country (see Head and Ries 2002; Greenaway and Nelson 2001; Xu 2000; Borensztein et al. 1998; Berman et al. 1998; Benhabib and Spiegel 1994; Nelson and Phelps 1966, etc., for surveys) and on the other, MNCs influence human capital through spillovers (see Blomström and Kokko 2003; Slaughter 2002; Gorg and Strobl 2001, 2003; Kokko et al. 2001, etc., for related discussion). Zhang and Markusen (1999); Lucas (1990) and Dunning (1988) observed that cheap labour may not be sufficient for attracting FDI, quality of the labour force is also crucial. Noorbakhsh et al. (2001) find the levels of human capital as one of the most important determinants of attracting FDI. Improvements in education and level of human capital increase the absorption capacity of foreign technology by the domestic firm. This view suits well for countries like Korea and Taiwan where a large amount of FDI flowed only after the countries were able to develop a workforce that was highly educated. This line of argument suggests that countries with high initial endowment of human capital attract foreign capital that subsequently brings about skill-biased adjustments in sectoral capital–labour ratios. Xu (2000) uses a multicountry study of technology diffusion effect of multinational enterprises to show that for the LDCs US affiliates have positive productivity effect on the host country but the author found no evidence of relating positive productivity to technology transfer.[6] Regression results of the model shows that positive effects depend on the threshold level of human capital which lies somewhere between 1.4 and 2.4 years of male secondary school attainment which was much higher than 0.52 as estimated by Borensztein et al. (1998).[7] Most of this literature points on the existing 'skill gap' that MNCs face while investing in developing countries. However, the empirical studies suggest that this 'skill gap' is endogenous with FDI. Miyamoto (2003) emphasized that 'skill gap' and FDI reinforce each other through complementary channels. Kar and Sinha (2014) in a theoretical study develop an aggregate transmission mechanism to show that technology deepening through MNC activities in the advanced sectors affect economywide skill formation. Accordingly, deepening of technology unambiguously raises the aggregate skill formation of the economy when the MNCs are more skill intensive than the traditional sector of production.

viii. Institutional Quality

In the recent years, institutional quality of the developing countries has qualified as one of the most important factors to attract FDI (see, Bissoon 2011; Ali et al. 2010; Kinda 2010; Wernick et al. 2009; Busse and Hefeker 2007; Daude and Stein 2007; Anghel 2005; Stein and Daude 2001; Hausman and Fernandez-Arias 2000; Wei 1997, 2000; Shleifer and Vishny 1993; Wheeler and Mody 1992). Wheeler and Mody (1992) made an early attempt to investigate the role of good institutions on FDI by taking a composite measure of risk factors that included institutional variables like extent of bureaucratic red tape, political instability, corruption and quality of the legal system. These factors were then tied up together with other factors such as, attitudes towards private sector, living environment, inequality, risk of terrorism, etc., making it complex to identify the factors separately in the index. However, taking the first principal component of 13 risk factors of their composed index, Wheeler and Mody (1992) did not find any significant impact of good quality institutions on the location of US multinationals. Later studies done by Wei (1997, 2000) found negative association between corruption and FDI. Using a broader range of institutional variables, Stein and Daude (2001) showed that quality of institutions plays significantly in attracting FDI. Hausman and Fernandez-Arias (2000) found better institutions to play negative role in attracting FDI. Accordingly, foreign capital inflows other than FDI were much more sensitive to good

[6]The reason was that LDCs were not endowed with sufficient amount of human capital to absorb the technology diffusion of the MNEs.

[7]"This threshold value is much higher than the 0.52 years estimated by BGL (1998). BGL's estimate is the human capital threshold to benefit from the presence of MNEs, while our estimate is the human capital threshold to benefit from technology transfer of MNEs. Most LDCs meet the first threshold but not the second. Our results are consistent with the findings of the previously mentioned single-country studies; technology spillover effects of MNEs are positive and significant in advanced countries but are insignificant in less developed countries" (Xu 2000, p. 479).

institutional quality. In another study, Globerman and Shapiro (2002) found good governance to have positive influence both on inflows and outflows of FDI, however, the influence on outflows was significant only for developed countries. Zheng (2006) found inverted U-shape relationship between institutional quality and FDI flows. Accordingly, strengthening of institutions simply created a downfall in the inflows of FDI after a period of high flows associated with low-quality institution. Intuitions may be that strengthening of political institutions was guided by new rules that did not suite the foreign investors.

4.2 **Push** *Versus* **Pull** *Factors as Determinant of FDI Inflows*

Importance of *push* versus *pull* factors in driving FDI has been discussed in the works of Calvo et al. (1993), Chuhan et al. (1993), Schadler et al. (1993), Hernandez and Rudolph (1994), Dooley, Fernandez-Arias, and Kletzer (1996) and Fernandez-Arias (1996). In Calvo et al. (1993) foreign exchange reserves were taken as a proxy to capital inflows in Latin America; Fernandez-Arias (1996) measured quarterly portfolio capital inflows for a panel of middle-income developing countries and the study by Dooley, Fernandez-Arias and Kletzer (1996) took price of commercial bank debt as a proxy of capital inflows. All these three studies found external factor to be the prime determinants of capital inflows in spite of their modelling differences. Dooley, Fernandez-Arias and Kletzer (1996) showed external factors to play the leading role in explaining rise in prices leaving no role for the domestic environment.

Foreign capital inflows to Latin America during the second half of 1980s that increased significantly during 1990 and 1991 was partly explained by external conditions such as continuing recession, falling of world interest rate and balance of payment developments in the US (Calvo et al. 1993). Calvo et al. (1993) used principal component analysis to show significant co-movement among foreign reserves and real exchange rates for ten Latin American countries for the time period 1990–1991. Structural VARs conducted in the model shows that foreign factors contribute significantly in accounting for movement in reserves and real exchange rate. According to Fernandez-Arias (1996), linkages and channels in his study have been discussed analytically and empirically, unlike the study of Calvo et al. (1993) where conclusions relied only on statistical analysis of common factors. In addition according to Fernandez-Arias (1996), findings of Calvo et al. (1993) were not in line with traditional portfolio models that are guided by return differentials. Importance of external factors supported in the above-mentioned works has been challenged by Schadler et al. (1993), where the authors have identified that in many cases timing of the change in external factors do not match the timing of internal flows. Moreover, country-specific factors played a huge role that was ignored by the above-mentioned studies. On the other hand, Hernandez and Rudolph (1994) have shown how careful specification of domestic factors could explain long-term credit flows for a sample of 22 developing countries for the time period 1986–1993. They found statistical significance of domestic credit worthiness with no role of external factor.

4.3 Business Cycles in Advanced Economies and Flow of FDI to Emerging Markets

Cycles of economic development in United States are felt in many of the developed countries, and the same holds true for many emerging economies of Asia, Eastern Europe and Latin America. Central banks opt for easy monetary policy and lower down the interest rate to dampen the effect of business cycle during the recession and hike the interest rate with the signs of boom. Falling of US interest rate and accumulation of foreign exchange reserves in the banks of Latin American countries and associated currency appreciation highlighted in the works of Calvo et al. (1993) have been supported by subsequent studies. In another study conducted by Reinhart and Reinhart (2001) over the period 1970–1999, it was shown that FDI, which is more stable than other forms of capital outflows, shrank during US recession and short-term capital continued to grow. More specifically, outflow of FDI and portfolio capital to the emerging economies increased when there were economic expansions in the US coupled with falling interest rate (Reinhart and Reinhart 2001). On the contrary, other flows (bank lending) increased with US recession and falling interest rate. This disparity between FDI and other forms of capital flows owes to the lending operation of the banks abroad in the midst of falling domestic demand and declining interest rate during the recession. However, the study found that regionwise the composition of capital flows were different for different parts of the business cycle. Rising Asia, Middle East and Europe showed signs of slowing down with downturn in US economy, while for Africa and Western Hemisphere the picture was different (Reinhart and Reinhart 2001). Works done by Fernandez-Arias (1996), Frankel and Roubini (2000) and Kaminsky and Schmukler (2001) show evidence that as the cost of borrowing international capital falls more than fall in the international interest rate for many emerging countries, their country-risk premia moves with international interest rate in a manner that magnifies the interest rate cycles of the industrialized countries (Reinhart and Reinhart 2001). Most recently, Albuquerque et al. (2005) has shown negative relationship between rising interest rate in the advanced countries and outflow of FDI to the developing countries.

Studies of Calvo et al. (1993) and Reinhart and Reinhart (2001) focused on aggregate data on FDI flows and used US cycle as a proxy for source country cycle though US only accounted for nearly 30 % of the total outflow and inflow of FDI from the perspective of OECD countries. These limitations have been addressed in the work of Yeyati et al. (2007) where bilateral FDI flows for 22 OECD source countries are taken and the cycles of both the source and host countries are identified separately. Findings of the paper confirm the conclusions of the previous studies. Interest rate cycles of the source countries again came out to be an important determinant of FDI flows. For Europe and US, outflow of FDI was countercyclical with the business cycle (for Japan the opposite was true) and the outflows declined with rising interest rate (Yeyati et al. 2007). However, the overall contribution of business cycles of the industrialized countries in outward FDI movement is not unanimous and eventually depends on empirical issues. In another paper, Norris et al. (2010) uses data on bilateral FDI outflows from G7 countries to low-income countries to document the role of economic conditions of the developed countries in explaining cross-country variations in FDI inflows in the recent period. Accordingly, economic conditions of the developed countries are found to be significant determinant of cross-country variations of FDI inflows in the recent years.

4.4 Internalization by the Global South

Though outward investment by the emerging economies is not a new phenomenon, the past couple of decades have witnessed a surge in the quantity and also qualitative transformation in the pattern of their investment. There has been a significant increase in the outward investment by the developing country multinationals since 1980. The figures increased from $80 billion in 1980 to $129 billion in 1990 to more than $1 trillion in 2004 (UNCTAD 2004). The emergence of second wave of MNEs was quite different from the *first wave* or the *pre-globalization* success stories that were mainly driven by domestic *push* factors such as market restrictions and export difficulties. According to Yeung (2000), the rise of *second-wave* MNEs from emerging market economies, best known as *latecomers* "*is less driven by cost factors* per se*, but more by a search for markets and technological innovations to compete successfully in the global economy*". The *latecomer firms* utilized these *pull* factors for their rapid internalization. However, the emergence of this *second-wave* firms was a paradox as their sudden appearance cannot be explained by conventional strategies adopted by MNCs.

It has been well recognized in the literature that flow of outward FDI has been a result of interaction between domestic *push* factors, e.g. internal policies, inadequateness of domestic market (in terms of scale and opportunities to expand), tough competition at home, export difficulties and external *pull* factors. There are a number of studies (see Aykut and Ratha 2004) that attempt to explain the relative importance of these factors that drive outward FDI from the developing countries.

Below we list the different factors responsible for outward FDI from developing countries and also highlight some of the respective studies dealing with them.

4.4.1 Drivers of Outward FDI for the Global South: Empirical Evidence

One of the most important domestic *push* factors that drive out FDI from developing countries is *market and trade related conditions*. According to UNCTAD (2006), internalization by the Chinese firms is mainly driven by poor domestic market opportunities and trade barriers. Moreover, rising *labour cost* has become a cause of concern for the MNCs from countries such as Malaysia, Republic of Korea, Singapore and Mauritius (see Schive and Chen 2004; Brooks and Mirza 2005). Inflationary pressure has been one of the main driving factors for countries like, India, China and Turkey during the 1990s (see Erdilek 2005; Banga 2006). According to Hymer (1976) and Athreye and Kapur (2009) firms often rush to invest abroad to increase their *competitiveness* vis-a-vis their domestic rivals (e.g. strong competitive element has been found in the overseas investment by the Indian MNCs). Moreover, pressure mounting due to competition from *low-cost production* (mainly from East and South-East Asian manufactures) in the international market has been the cause of internalization for many Latin American and African countries (see ECLAC 2006; Farrell et al. 2005; Gaulier et al. 2006). Moreover, in an integrated world, *opportunities of competition* from foreign companies have become another major source of internalisation by the firms from the developing economies. According to Nolan (2001), Jürgens and Rehbehn (2006), rapid increase in Chinese outward FDI is mainly attributed to competition from foreign companies. However, competition abroad also provides incentive for firms from developing countries to invest abroad. It has been shown by Fortanier and van Tulder (2009), in some sectors (e.g. chemicals and pharmaceuticals, oil and petroleum and telecommunications) MNCs from developing countries set up joint ventures with the existing MNCs from the developed countries to compete globally. In addition, home country *government policies* and *adverse business conditions* also play significant roles in outward FDI. Transparent governance, investment in infrastructure, property rights, minimal exchange rate regulations and other macroeconomic conditions of the home country determine the location choice of FDI. According to a survey conducted by UNCTAD, decision of the Chinese firms going global was mainly guided by home *government policies*. Moreover, *adverse business conditions* mainly generated from inadequate infrastructure, labour issues, or undeveloped input or component services push firms to invest in foreign countries (e.g. labour issues have played a significant role in South Africa in limiting domestic investment and in possible rise in foreign investment).

However, merely studying of *push* and *pull* factors may not be sufficient to understand the ultimate location choice of the developing country MNCs; understanding of their motives and strategies is also essential (UNCTAD 2006). A particular driver that affects different MNCs may lead to different motives and strategies thus ending up with different location choices by the firms. For example, competition faced in the domestic sphere may lead a firm to invest overseas, but it can respond to this pressure in a number of ways. For instance, the firm may go for searching new customers in a middle-income developing country by taking *market-*

seeking attitude; or it may go for *efficiency-seeking* by investing in a lower income developing country to lower its cost of production; or it may adopt *resource-seeking* behaviour in search for key inputs may be in a country with abundant supply of raw materials; or even it may go for *created asset-seeking* in developed economies; or may go for a *mixed strategy*. Therefore, whereas, *resource-seeking*, *market-seeking* and *efficiency-seeking* motives are to be found as the dominant factors driving out FDI to other developing countries; *strategic asset-seeking* has been the main motivation for FDI flying from developing countries to developed countries.

Below we discuss different motives of MNCs that lead to location choice of outward FDI from the developing economies.

4.4.2 Motivations

Coming to the *market-seeking* motive, accessing the global market has been one of the most important motives of outward FDI from the developing countries (see Aykut and Goldstein 2006; Athreye and Kapur 2009). Though evidence of *market-seeking* behaviour is very common in most of the industries, variations occur depending on the nature of the source country (e.g. outward FDI from South Africa is very common in industries such as chemicals, food and beverages, finance, and transport and communication). Though, developed country markets are very attractive due to their large market size and accessibility due to regional integration, especially, North America and Europe; theory and evidence suggest that most of the MNCs from developing countries, such as, Latin America, Africa, East and South-East Asia, initially invest in their neighbouring economies especially due to familiarity and common factors. Recently, market size has generated increasing interest in South-South investment and trade corridors by some of the Brazilian, Indian, Chinese and South African firms seeing prospects in each other's relatively large markets.[8] However, according to UNCTAD (2006) greater relative importance is given to developed country markets than to developing country markets by a majority of firms from India, China, Korea and Russia. Moreover, apart from capturing the vast markets of the developed economies through exports, some of the affiliates were established to get proximity to their clients (for example Indian IT firms) and some to get access to the foreign markets against protectionist barriers (investment by Chinese firms).

Another motivation for the firms from the developing countries (mainly countries that lack natural resources, e.g. India, China and Turkey) to invest abroad is to get access to natural resources; especially raw materials (see Makino et al. 2002; Ariff and Pio Lopez 2007; Cuervo-Cazurra 2007; Buckley et al. 2007; Kumar and Chadha 2009 etc.). Buckley et al. (2007) found *resource-seeking* from developing countries to be an important motivation for Chinese outward FDI. In another study,

[8] See Kaplinsky and Morris (2006), Naidu (2005), Rios-Morales and Brennan (2006), Goldstein and Toulan (2005) for literature on trade corridors.

Kumar and Chadha (2009) find that most of the *resource-seeking* outward FDI from Chinese and Indian firms are driven by the motive to secure the supply of raw materials for the development of their respective homelands. State-owned enterprises namely, China National Petrol Corporation (CNPC), China National Offshore Oil Corporation (CNOOC), India's Oil and Natural Gas Corporation (ONGC), Turkish Petroleum Corporation (TPC), etc., are typical firms in this category. Most of these firms from countries with poor reserve of natural resources have invested in the areas determined not by regional proximity but by availability of resources. For example, ONGC from India has a vast area of operation that includes Algeria, Brazil, Côte d'Ivoire, Cuba, the Islamic Republic of Iran, Kazakhstan, Nepal, Nigeria, Qatar, the Russian Federation, Syrian Arab Republic, Sudan and Venezuela. In another study, Goldstein (2008) showed that outward FDI activities of Tata Chemicals and Tata Power have been guided by their access to natural resources required for their production process.

The *efficiency-seeking* outward FDI is an important purpose for companies mostly from Asia and for industries, namely, electrical and electronic products, garments and footwear and IT services (see Ariff and Pio Lopez 2007; Sim and Pandian 2007; Kazmi 2006; Chen and Lin 2005; Page and Velde 2004; Lim 2005; Moon 2005; Zainal 2005; Cherry 2001, etc.). MNCs for whom *efficiency-seeking* motive is very important mainly comes from Hong Kong (China), Malaysia, Mauritius, the Republic of Korea and Taiwan Province of China. Most of these countries nowadays face relatively high cost of labour that has forced them to invest in lower cost locations that has generated, in some cases, regional integrated production system (Samsung is a typical example as it has production facilities all over South-East Asia). For firms from Taiwan province of China efficiency means low-cost production and for the Indian firms efficiency represents synergies obtained from international integration in production and services. However, there are a few firms from China and Singapore that go for low-cost *efficiency-seeking* outward FDI.

There are some studies like Buckley et al. (2007); Cross and Voss (2008); Liu and Tian (2008) that focus on *strategic asset-seeking* motive together with *market-seeking*. Firms adopt *mixed motives* when they invest for more than one purpose simultaneously. Singapore Technologies Telemedia has ventured in a number of markets including Indonesia, United Kingdom, United States, and in many Latin American countries since 2002. According to UNCTAD (2006) all of its foreign affiliates are established to access the local market and at the same time secure strategic assets and create synergies. *Complementary motives* are found in many firms when they combine more than one motive to pursue their goal. Integrated Microelectronics Inc (IMI), a Philippines-based company has gone for *complementary motives* when it adopted two strategies such as *created asset-seeking* and *market-seeking* by acquiring its first foreign affiliate in US to improve its own R&D and at the same time buying a Singapore-based affiliate with manufacturing facilities in Singapore and China to improve its competitive position in China's electronics market. Regarding the *evolutionary motives* Pradhan (2007) has identified the evolution of motivations of the Indian MNCs that started mainly with *market-*

seeking activities in the pre-liberalization period aiming towards other developing countries. This strategy shifted to *resource-seeking* and more recently to *strategic asset-seeking* in developed countries.

5 Concluding Remarks

Increase in capital movement in different parts of Asia, Eastern Europe, Latin America, the Caribbean and the Russian Federation in the past couple of decades suggests that both the global factors and the favourable domestic conditions of the developing countries played important roles in driving FDI flows to these countries. The rush in global inflows to the middle-income and low-income countries before the global financial crisis of 2008 occurred in the strong background of global economic growth and strong terms of trade. Emerging economies continued to grow at a high rate with increasing South-South trade and FDI linkages. Moreover, favourable financial conditions accompanied by low world interest rates resulted in huge global liquidity and low borrowing costs. Though most of the related literature in this regard has focused on the relative importance of *push* and *pull* factors to drive global capital to the new set of countries, there is another strand of thought that focuses on the macroeconomic *countercyclical* policies adopted by the capital importing countries. According to World Bank, Latin America and the Caribbean were the least affected in the 2009 global economic meltdown compared to other regions. Accordingly, whereas weak currencies, fiscal processes and banking system magnified the degree of previous crisis in the region, improved macroeconomic and financial conditions helped the economy to stay balanced during 2009. In this case, effective implementation of *countercyclical* policies not only supported domestic demand for the large countries of the region but has also met the rising demand from fast growing emerging markets, e.g. China.

Moreover, business cycles in the industrialized countries also play an important role in determining capital flows to the rising economies. Interest rate cycles coupled with *countercyclical* monetary policy of the developed countries influence emerging economies' access to international capital. Regarding the outflow of FDI from the Global South, though outward investment by the emerging economies is not a new phenomenon, the past couple of decades have witnessed a surge in the quantity and also qualitative transformation in the pattern of their investment. Survey of the literature finds that besides studying the *push* and *pull* factors it is important to study the motives and the strategies of the developing country MNCs to know their location choice. These firms may be guided by *market-seeking*, *resource-seeking* or *efficiency-seeking* behaviour while investing in other countries. Clearly, further research is needed to understand the determinants of FDI outflows from the Global South, and the standard theories of the determinants of FDI may need to be revisited, in light of this very recent phenomenon, that is sharply increasing in significance in global capital flows.

References

Albuquerque R, Loayza N, Serven L (2005) World market integration through the lens of foreign direct investors. J Int Econ 66(2):267–295

Ali F, Fiess N, MacDonald R (2010) Do institutions matter for foreign direct investment? Open Econ Rev 21(2):201–219

Anghel B (2005) Do institutions affect foreign direct investment? International doctorate in economic analysis, 2–40. ASEAN Bulletin (2010), ASEAN Secretariat

Anyanwu JC, Erhijakpor EO (2004) Trends and determinants of foreign direct investment in Africa. W Afr J Monetary Econ Integr 4(2):21–44

Ariff M, Pio Lopez G (2007) Outward foreign direct investment: the Malaysian experience, mimeo, Malaysian institute of economic research (MIER)

Artige L, Nicolini R (2006) Evidence on the determinants of foreign direct investment: The case of three European regions. CREPP working papers 0607

Asiedu E (2002) On the determinants of foreign direct investment to developing countries: Is Africa different? World Dev 30(1):107–119

Athreye S, Kapur S (2009) Introduction: the internationalization of Chinese and Indian firms—trends, motivations and strategy. Ind Corp Change 18(2):209–221

Aulakh PS (2007) Special issue on emerging multinationals from developing economies: motivations, paths and performance. J Int Manag 13(3):235–402

Aykut D, Goldstein A (2006) Developing country multinationals: South-South investment comes of age. OECD working papers no. 257

Aykut D, Ratha D (2004) South-South FDI flows: how big they are? Transnational Corporations 13(1):149–175

Bandera VN, White JT (1968) U.S. direct Investments and domestic markets in Europe. Economia Internazionale 21:117–133

Banga R (2006) What triggers outward foreign direct investment from Asian developing countries? Mimeo

Barrel R, Pain N (1998) Real exchange rates, agglomerations, and irreversibilities: macroeconomic policy and FDI in EMU. Oxford Rev Econ Policy 14(3):152–167

Benhabib J, Spiegel MM (1994) The role of human capital in economic development: Evidence from aggregate cross-country data. J Monetary Econ 34(2):143–173

Berman E, Bound J, Machin S (1998) Implications of skill-biased technological change: international evidence. Q J Econ 113(4):1245–1279

Bevan AA, Estrin S (2000) The determinants of FDI in transition economies. *Discussion Paper No.2638, Center for Economic Policy Research, London*

Billington N (1999) The Location of Foreign Direct Investment: An Empirical Analysis. Appl Econ 31(1):65–76

Bissoon O (2011) Can better institutions attract more foreign direct investment (Fdi)? evidence from developing countries. Int Conf Appl Econ–ICOAE 59

Blomström M, Kokko A (2003) Human capital and inward FDI, CEPR discussion paper no. 3762. London

Blonigen BA (1995) Explaining Japanese foreign direct investment in the United States. Unpublished Ph.D. dissertation, University of California, Davis

Blonigen BA (2005) A review of the empirical literature on FDI determinants. Atlantic Econ J 33 (4):383–403

Blonigen BA, Feenstra RC (1996) Effects of U. S. trade protection and promotion policies, NBER working papers no. 5285

Borensztein E, Gregorio J, Lee W (1998) How does foreign direct investment affect economic growth? J Int Econ 45(1):115–135

Brooks DH, Mirza HR (2005) Outward FDI from developing Asia. Paper presented at the Asian development bank seminar. In: Outward foreign direct investment from Asian developing countries, Bangkok, 28–29 Nov

Buckley P, Casson M (1976) The future of the multinational enterprise. Macmillan, London
Buckley PJ, Clegg J, Cross AR, Liu X, Voss H, Zheng P (2007) The determinants of Chinese outward foreign direct investment. J Int Bus Stud 38(4):499–518
Busse M, Hefeker C (2007) Political risk, institutions and foreign direct investment. Eur J Polit Econ 23(2):397–415
Calvo G, Leiderman L, Reinhart C (1993) Capital inflows to Latin America: the role of external factors. IMF staff papers no. 40
Cassou SP (1997) The link between tax rates and foreign direct investment. Appl Econ 29 (10):1295–1301
Chakrabarti A (2001) The determinants of foreign direct investment: sensitivity analyses of cross-country regressions. Kyklos 54:89–113
Chen EKY, Lin P (2005) Outward foreign direct investment from Hong Kong. Mimeo, Lingnan University, Hong Kong
Cherry J (2001) Korean multinationals in Europe. Curzon, Richmond
Chuhan P, Claessens S, Mamingi N (1993) Equity and bond flows to Latin America and Asia: the role global and country factors. World Bank, IECIF policy research working paper no. 1160
Cross AR, Voss H (2008) Chinese direct investment in the United Kingdom: an assessment of motivations and competitiveness, presented at Corporate Strategies in the New Asia, University of Bremen, 1–2 Feb 2008
Cuervo-Cazurra A (2007) Sequence of value-added activities in the multinationalization of developing country firms. J Int Manag 13(3):258–277
Culem CG (1988) The locational determinants of direct investments among industrialized countries. Eur Econ Rev 32:885–904
Daude C, Stein E (2007) Longitude matters: time zones and the location of foreign direct investment. J Int Econ 71(1):96–112
Dooley M, Fernandez-Ariaz E, Kletzer K (1996) Is the debt crisis history? recent private capital inflows to developing countries. World Bank Economic Review 10(1): 27–50
Dunning JH (1980) Toward an eclectic theory of international production: some empirical tests. J Int Bus Stud 11(1):9–31
Dunning JH (1981) International production and the multinational enterprise. Allen & Unwin, London
Dunning JH (1988) Explaining international production. Unwin Hyman, London
Dunning JH (2006) Towards a new paradigm of development: implications for the determinants of international business. Transnational Corporations 15(1):173–227
ECLAC (2006) Foreign investment in Latin America and the Caribbean 2005. United Nations, Santiago, Chile
Edwards S (1990) Capital flows, foreign direct investment, and debt-equity swaps in developing countries. NBER working paper no. 3497
Eitemann DK, Stinehill AI, Moffett MH (2007) Multinational business finance, 11th edn. Pearson Addison Wesley, Boston
Erdilek A (2005) Case study on outward foreign direct investment by enterprises from Turkey. Paper prepared for the UNCTAD expert meeting on enhancing the productive capacity of developing country firms through internationalization, Geneva, 5–7 Dec 2005
Farrell D, Puron A, Remes JK (2005) Beyond cheap labor: lessons for developing economies. McKinsey Quarterly 1:99–109
Fernandez-Arias E (1996) The new wave of private capital inflows: Push or pull? J Dev Econ 48:389–418
Flamm K (1984) The volatility of offshore investment. J Dev Econ 16(3):231–248
Fortanier F, van Tulder R (2009) Internationalization trajectories—a cross-country comparison: are large Chinese and Indian companies different? Ind Corp Change 18(2):223–247
Frankel JA, Roubini N (2000) The role of industrial country policies in emerging market crises. Paper prepared for the NBE conference on economic and financial crises in emerging market economies, Woodstock

Froot KA, Stein JC (1991) Exchange rates and foreign direct investment: an imperfect capital markets approach. Quart J Econ 106(4):1191–1217
Gastanaga VM, Nugent JB, Pashamova B (1998) Host country reforms and FDI inflows: how much difference do they make? World Dev 26(7):1299–1313
Gaulier G, Françoise L, Deniz Ü-K (2006) China's emergence and the reorganisation of trade flows in Asia. CEPII working paper no. 2006–05
Globerman S, Shapiro D (2002) Global foreign direct investment flows: the role of governance infrastructure. World Dev 30(11):1899–1919
Goldsbrough DG (1979) The role of foreign direct investment in the external adjustment process. IMF staff papers 26:725–754
Goldstein A (2008) The internationalization of Indian companies: the case of Tata. Centre for advanced study on India (CASI) working paper no. 08–02, University of Pennsylvania
Goldstein A, Toulan O (2005) Multlatinas' go to China: two case studies, mimeo
Gorg H, Strobl E (2001) Multinational companies and productivity spillovers: a meta-analysis. Econ J 111(475):F723–F739
Gorg H, Strobl E (2003) Multinational companies, technology spillovers and plant survival. Scand J Econ 105(4):581–595
Greenaway D, Nelson D (2001) Globalization and labour markets: literature review and synthesis. Research Paper No. 2001/29, Research paper series, globalisation and labour markets programme, University of Nottingham, Nottingham
Grubert H, Mutti J (1991) Taxes, tariffs and transfer pricing in multinational corporate decision making. Rev Econ Stat 73(2):285–293
Guisinger S (1985) Investment incentives and performance requirements: patterns of international trade, production, and investment. Praeger, New York
Gupta VK (1983) A simultaneous determination of structure, conduct and performance in Canadian manufacturing. Oxford Econ Pap 35(2):281–301
Hailu ZA (2010) Demand side factors affecting the inflow of foreign direct investment to African countries: does capital market matter? Int J Bus Manage 5(5):104–116
Hartman DG (1984) Tax policy and foreign direct investment in the United States. National Tax J 37(4):475–487
Hausman R, Fernandez-Arias E (2000) Foreign direct investment: good gholesterol? working paper no. 417, Inter-American Development Bank
Head K, Ries J (2002) Offshore production and skill upgrading by Japanese manufacturing firms. J Int Econ 58(1):81–105
Hennart J (1982) A theory of multinational enterprise. University of Michigan, Ann Arbor
Hernandez L, Rudolph H (1994) Domestic factors, sustainability, and soft landing in the new wave of private capital inflows, World Bank, Mimeo, Nov 1994
Hines JR, Rice EM (1994) Fiscal paradise: foreign tax havens and American business. Q J Econ 109:149–182
Hymer SH (1976) The international operations of national firms: a study of foreign direct investment. MIT Press, Cambridge, MA
Jackson S, Markowski S (1995) The attractiveness of countries to foreign direct investment. J World Trade 29:159–180
Jaspersen S, Aylward A, Knox A (2000) The effects of risks on private investment: Africa compared with other developing areas. In: Collier P, Pattillo C (eds) Investments and Risks in Africa. St. Martin's Press, New York, pp 71–95
Jürgens U, Rehbehn R (2006) China's changing role in industrial value chain and reverberations on industrial actors in Germany. In: Söderman Sten (ed) Emerging multiplicity: integrated and responsiveness in Asian business development. Palgrave MacMillan, Basingstoke
Jordaan JC (2004) Foreign direct investment and neighboring influences. Unpublished doctoral thesis, University of Pretoria
Kaminsky GL, Schmukler S (2001) Emerging market instability: do sovereign ratings affect country risk and stock returns? mimeograph, The World Bank, March 2001

Kaplinsky R, Morris M (2006) Dangling by a thread: how sharp are the Chinese scissors? Paper prepared for United Kingdom DFID trade division, Institute of Development Studies, Brighton
Kar S, Sinha C (2014) Sectoral technical progress and aggregate skill formation. J Ind Competition Trade 14(2):159–172
Kazmi A (2006) Motives for enterprise internationalization of Malaysian transnational corporations. Paper presented at the annual conference of the academy of international business (United Kingdom Chapter), Manchester, 7–8 April
Kemsley D (1998) The effect of taxes on production location. J Account Res 36:321–341
Kinda T (2010) I nvestment climate and FDI in developing countries: firm-level evidence. World Dev 38(4):153–498
Kindlerberger C (1969) American business abroad: six lectures on direct investment. Yale University Press, New Haven
Kokko A, Zejan M, Tansini R (2001) Trade regimes and spillover effects of FDI: evidence from Uruguay. Rev World Econ 137(1):124–149
Kravis IB, Lipsey RE (1982) The location of overseas production and production for export by U. S. Multinational Firms. J Int Econ 12(3–4):201–223
Kumar N, Chadha A (2009) India's outward foreign direct investments in steel industry in a Chinese comparative perspective. Ind Corp Change 18(2):249–267
Leitão NC (2010) Localization factors and inward foreign direct investment in Greece. Theor Appl Econ 17 (6(547)):17–26
Lim WC (2005) Top glove corporation berhad. Paper prepared for the UNCTAD expert meeting on enhancing the productive capacity of developing country firms through internationalization, Geneva, 5–7 Dec
Lipsey RE (1999) The location and characteristics of US affiliates in Asia. NBER working paper no. 6876
Liu L, Tian Y (2008) The internationalisation of Chinese enterprises: the analysis of the UK case. Int J Technol Globalization 4(1):87–102
Loree D, Guisinger SE (1995) Policy and non-policy determinants of U.S. equity foreign direct investment. J Int Bus Stud 26(2):281–300
Lucas REB (1990) Why doesn't capital flow from rich to poor countries? Am Econ Rev 80(2):92–96
Makino S, Lau CM, Yeh RS (2002) Asset-exploitation versus asset seeking: implication for location choice of foreign direct investment from newly industrialized economies. J Int Bus Stud 33(3):403–421
Masayuki H, Razafimahefa I (2005), The determinants of foreign direct investment into Japan. Kobe University Economic Review 51
Mathews JA (2002a) Dragon multinationals—a new model for global growth. Oxford University Press, Oxford
Mathews JA (2002b) Competitive advantages of the latecomer firm: a resource-based account of industrial catch-up strategies. Asia Pac J Manag 19:467–488
Mathews JA (2006) Dragon multinationals: new players in 21st century globalization. Asia Pac J Manag 23(1):5–27
Miyamoto K (2003) Human capital formation and forign direct investment in developing countries. OECD working paper no. 211
Moon H-C (2005) Outward foreign direct investment by Korean firms. Paper prepared for the UNCTAD expert meeting on enhancing the productive capacity of developing country firms through internationalization, geneva, 5–7 Dec
Morisset J (2000) Foreign direct investment in Africa-Policies also matter. Transnational Corporation 9(2):107–125
Naidu S (2005) South Africa's relations with the People's Republic of China: mutual opportunities orhidden threats? In: Buhlungu Sakhela, Daniel John, Southall Roger, Lutchman Jessica (eds) State of the Nation: South Africa 2005–2006. HSRC Press, Cape Town, pp 457–483

Nankani GT (1979) The intercountry distribution of direct foreign investment. Garland, New York
Narula R (2006) Globalization, new ecologies, new zoologies, and the purported death of the eclectic paradigm. Asia Pac J Manag 23(2):143–151
Narula R, Wakelin K (2001) The pattern and determinants of U.S foreign direct investment in industrialised countries. In: Narula R (ed) Trade and Investment in a globalising world. Pergamon, New York, p 2001
Nelson RR, Phelps ES (1966) Investment in humans, technological diffusion and economic growth. Am Econ Rev 56(1/2):69–75
Nolan P (2001) China and the global economy. Palgrave, Basingstoke
Nonnenberg MB, Cardoso de Mendonça MJ (2004) The determinants of foreign investment in developing countries. http://ssrn.com/abstract=525462
Noorbakhsh F, Paloni A, Youssef A (2001) Human capital and FDI inflows to developing countries. World Dev 29(9):1593–1610
Norris E, Honda J, Lahreche A, Verdier G (2010) FDI flows to low income countries: global drivers and growth implications. IMF working paper no. WP/10/132
Owen RF (1982) Inter-industry determinants of foreign direct investments: a Canadian perspective. In: Rugman AM (ed) New theories of multinational enterprise. Croom Helm, London
Page S, Velde D (2004) Foreign direct investment by Africa countries. Paper prepared for In Went/UNCTAD meeting on FDI in Africa, Addis Ababa, 22–24 Nov 2004
Papanastassiou M, Pearce R D (1990) Host country characteristics and the sourcing behaviour of UK manufacturing industry. University of Reading, Department of Economics, Discussion Papers in International Investment and Business Studies, Series B, 2, No. 140.
Pistoresi B (2000) Investimenti diretti esteri e fattori di localizzazione: L'America Latina e il Sud Est asiatico. Rivista di Politica Economica 90:27–44
Porcano TM, Price CE (1996) The effects of government tax and non-tax incentives on foreign direct investment. Multinational Bus Rev 4(1):9–20
Pradhan JP (2007) Growth of Indian multinationals in the world economy. Implications for development. ISID working paper no. 2007/04
Quazi RM (2007) Investment climate and foreign direct investment: a study of selected countries in Latin America. Glob J Bus Res 1(2):1–13
Reinhart C, Montiel P (2001) The dynamics of capital movements to emerging economies during the 1990s. MPRA paper no. 7577, Munich Personal RePEc Archive, Munich, Germany
Reinhart C, Reinhart V (2001) What hurts most? G-3 exchange rate volatility or interest rate volatility. NBER working paper no. 8535
Resmini L (2000) The determinants of Foreign direct investment in the CEECs. New evidence from sectoral patterns. Econ Transit 8(3):665–689
Sader F (1993) Privatization and foreign investment in the developing world. World bank working paper no. 1202
Saunders RS (1982) The determinants of foreign direct investment. Can J Econ 15:77–84
Schadler S, Carkovic M, Bennett A, Kahn R (1993) Recent experiences with surges in capital inflows. Occasional paper 108, International Monetary Fund, Washington
Schive C, Chen T (2004) The globalization of business in Taiwan. In: Chen Tain-Jy, Lee Joseph S (eds) The New knowledge economy of Taiwan. Edward Elgar, Cheltenham
Schmitz A, Bieri J (1972) EEC tariffs and U.S. direct investment. Eur Econ Rev 3(3):259–270
Schneider F, Frey B (1985) Economic and political determinants of foreign direct investment. World Dev 13(2):161–175
Sekkat K, Veganzones-Varoudakis M-A (2007) Openness, investment climate, and FDI in developing countries. Rev Dev Econ 11(4):607–620
Shamsuddin AF (1994) Economic determinants of foreign direct investment in less developed countries. Pak Dev Rev 33(1):41–51
Shapiro CA (2006) Multinational financial management, 8th edn. Wiley, United States of America
Shleifer A, Vishny R (1993) Corruption. Q J Econo 108(3):599–617

Sim AB, Pandian JR (2007) An exploratory study of internationalization strategies of Malaysian and Taiwanese Firms. Int J Emer Markets 2(3):252–273
Singh H, Jun K (1995) Some new evidence on determinants of foreign direct investment in developing countries. Policy research working paper no: 1531
Slaughter M (2002) Skill upgrading in developing countries: has inward foreign direct investment played a role? OECD working paper no. 192
Stein E, Daude C (2001) Institutions, integration and the location of foreign investment. Inter-American Development bank, Research department, Mimeographed document
Swedenborg B (1979) The multinational operations of Swedish firms: analysis of determinants and effects. Working paper, Industrial Institute of Economic and Social Research, Stockholm
Swenson DL (1994) The impact of U. S. tax reform on foreign direct investment in the United States. J Pub Econ 54(2):243–266
Teece D, Pisano G, Suen A (1997) Dynamic capabilities and strategic management. Strateg Manag J 18(7):509–533
Tsai P (1994) Determinants of foreign direct investment and its impact on economic growth. J Econ Dev 19(1):137–163
Tuman JP, Emmert CF (1999) Explaining Japanese foreign direct investment in Latin America, 1979–1992. Soc Sci Q 80(3):539–555
UNCTAD (2004) World investment report 2004, United Nations
UNCTAD (2006) World investment report (2006), United Nations
UNCTAD (2014) World investment report (2014), United Nations
Vernon R (1966) International investment and international trade in the product cycle. Quart J Econ 80(2):190–207
Wafure GO, Nurudeen A (2010) Determinants of foreign direct investment in Nigeria: an empirical analysis. Global J Hum Soc Sci (1):26–35
Wei S-J (1997) How taxing is corruption on international investors? NBER working paper no. 6030
Wei S-J (2000) How taxing is corruption on internal investors? Rev Econ Stat 82(1):1–11
Wernick DA, Haar J, Singh S (2009) Do governing institutions affect foreign direct investment? Int J Econ Bus Res 1(3):317–322
Wheeler D, Mody A (1992) International investment location decisions: The case with endogenous probability of reversal. Centre for the Study of African economies working paper 98–13, University of Oxford
Xu B (2000) Multinational enterprises, technology diffusion, and host country productivity growth. J Dev Econ 62(2):477–493
Yeung HWC (ed) (2000) The globalization of business firms from emerging economies. Edward Elgar, Cheltenham, UK
Yeyati L, Panizza U, Stein E (2007) The cyclical nature of North–South FDI flows. J Int Money Finance 26(1):104–130
Yih Yun YJ, Groenewold N, Tcha M (2000) The determinants of foreign direct investment in Australia. Econ Rec 76(232):45–54
Yulin N, Reed MR (1995) Locational determinants of U. S. direct foreign investment in food and kindered products. Agrobusiness 11(1):77–86
Zainal AY (2005) Outward foreign direct investment by Malaysian enterprises. Paper prepared for the UNCTAD expert meeting on enhancing the productive capacity of developing country firms through internationalization, Geneva, 5–7 Dec
Zhang KH, Markusen JR (1999) Vertical multinationals and host-country characteristics. J Dev Econ 59(2):233–252
Zheng Y (2006) Credibility versus flexibility political institutions and foreign direct investment. Mimeo, Annual Meeting of the American Political Science Association, August 30–September 3, Philadelphia, PA. World Investment Report 2010: Investing in a Low-carbon Economy (WIR10), UNCTAD

Foreign Direct Investment, Capital Formation, and Growth

Prabirjit Sarkar

Abstract This chapter examines the long-term effect of FDI on growth and capital accumulation in a panel of 65 countries covering a time-period of 1981–2005. It uses the dynamic panel data technique that considers the short-term relationships and the stable adjustment paths leading to the long-term relationships. It aggregates the individual country estimates through the alternative procedures such as mean group, pooled mean group, and dynamic fixed effect. It observes positive long-term effects of FDI on capital accumulation. There is some doubt regarding the impact of FDI on growth.

Keywords Foreign direct investment · Financial globalization · Liberalization

JEL Codes F21 · F43 · O10 · O16 · O50

1 Introduction

Freer flow of goods and services and freer mobility of capital are the order of the present era of globalization. Most of the mainstream economists are supporting free trade but they are divided on the issue of full-fledged financial globalization, which implies free movements of short-term and long-term capital. Bhagwati (1998), for instance, supported free trade but could not support full-fledged financial globalization because of its inherent crisis-prone nature. He argued that the 'Asian crisis cannot be separated from the excessive borrowings of short-term capital'. Before Bhagwati (1998), Stiglitz (1994) also criticized financial globalization on the grounds that financial markets are prone to market failures. Stiglitz (2000) argued that far more relevant (than other types of financial flows) for the long-term success of the economy is foreign direct investment (FDI).

P. Sarkar (✉)
Department of Economics, Jadavpur University, Kolkata, India
e-mail: prabirjit@gmail.com

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_17

There is some qualified empirical support for growth-promoting effect of FDI. Balasubramanyam et al. (1996) used cross-section data relating to a sample of 46 countries and supported the hypothesis of Bhagwati: the growth-enhancing effect of FDI is stronger in those countries which pursue an outward-oriented trade policy than it is in those countries following an inward-oriented policy.

Borensztein et al. (1998) studied the effect of FDI on economic growth in a cross-country regression framework; they utilized data on FDI flows from developed countries (DCs) to 69 less developed countries (LDCs) for the two decades, the 1970s and 1980s. Their findings suggest that 'FDI is an important vehicle for the transfer of technology, contributing relatively more to growth than domestic investment'. They also observed that FDI has the effect of increasing domestic investment suggesting a complementary relationship. There is a caveat: 'the higher productivity of FDI holds only when the host country has a minimum threshold stock of human capital'.

Time series and panel data analysis of De Mello (1999) for a sample of OECD and non-OECD countries in the period 1970–1990 arrived at the same conclusion as in Borensztein et al. (1998). He also has the caveat: 'the extent to which FDI is growth-enhancing depends on the degree of complementarity and substitution between FDI and domestic investment'.

The panel data analysis of Bengoa and Sanchez-Robles (2003) for a sample of 18 Latin American countries for 1970–1999 shows that FDI is positively correlated with economic growth in the host countries. 'The host country requires, however, adequate human capital, economic stability and liberalized markets to benefit from long-term capital flows' (Bengoa and Sanchez-Robles 2003, p. 529). Recently Li and Liu (2005) made a panel data analysis for 84 countries over the period 1970–99 and observed that the interaction of FDI with human capital exerted a strong positive effect on economic growth while that of FDI with the technological gap had a significant negative impact (see also Bende-Nabende et al. 2003).

Lipsey (2000), however, had a different view. He discounted the importance of FDI in capital formation: 'financing capital formation is not a primary role of FDI'. It is usually believed that capital moves in search of higher rates of return because of higher productivity of capital (due to capital scarcity) and/or lower labor, and raw material costs. Lipsey, however, points out 'the fact that most direct investment takes place among developed countries, where the differences in rates of return and capital abundance are not very great.' He observes inward and outward foreign direct investment (FDI) stocks and flows to go together, across countries and over time. The countries that invest a lot in this form tend also to be large recipients of FDI. Since much of FDI consists of offsetting two-way flows, he suggests that the main role FDI plays is that of transferring assets from less efficient to more efficient owners. He provides some evidence from the data on FDI flows between the United States and the rest of the world and comments that FDI can be viewed in recipient countries as freeing capital frozen in industries that the owners (including governments) would like to leave.

Nunnenkamp (2004) commented that 'the currently prevailing euphoria about FDI rests on weak empirical foundations' and showed that it is much more difficult for poor LDCs to derive macroeconomic benefits from FDI than to attract FDI. Durham (2004) analyzed data for 80 countries from 1979 to 1998 and observed that foreign portfolio investment as well as foreign direct investment does not have direct positive effects on growth; effects are contingent on the 'absorptive capacity' of host countries.

Some studies question the direction of causation from FDI to growth and argue that economic growth robustly causes growth in domestic investment and/or FDI (Choe 2003; Chakraborty and Basu 2002; Gao 2005; see also Blomstrom et al. 1996). There is one study of 73 DCs and LDCs over the period 1995–1999; it points out that corruption is a stimulus for FDI as corruption can be beneficial in circumventing regulatory and administrative restrictions (Egger and Winner 2005).

In this perspective the present study examines the relationship between FDI and growth in the less developed countries for which the relevant data are readily available. Our findings are presented in the next section (Sect. 2). In the last section concluding observations are made (Sect. 3).

2 Growth of FDI and Its Long-Term Relation with Economic Growth

Foreign Direct Investment (FDI) registered a rapid growth as shown in the data presented in Table 1. From 13 billion US dollar in 1970, total FDI flows rose to 54 billion in 1980 and further to 207 billion in 1990, 341 in 1995, and 1398 in 2000. During 2001–2003, it showed a tendency to decline; thereafter it started recovering. By 2006, it reached 1411 billion and rose further to 1833. In view of credit crunch of 2008, it declined. The share of DCs rose from 71 % in 1970 to 86 % in 1980; thereafter it showed a tendency to decline reaching 65 % in 1995 implying a rise in the share of LDCs in the rapidly rising FDI flows. The lion's share of the FDI flows to LDC went to East and South East Asia (65 % of LDC total and 22 % of the world total). Due to the Asian financial crisis, the share fell in 2000 (to 56 % of LDC total and 10 % of the world total) and the share of DCs rose. During the recovery stage, the share of Asia started picking up while the share of DC declined and came closer to the 1995 level (64–68 % of the world total).

The question is: does a country with a higher FDI (net) relative to its gross domestic product (GDP), FDIGDP, grow at a faster rate? We seek an answer to this question on the basis of World Bank data on World Development Indicators (WDI). Apart from FDIGDP we consider another variable, trade openness as measured by the share of trade (export plus import) in GDP, TRDGDP, explaining the growth of GDP per capita (PCYG). On the basis of the availability of data we have chosen a sample of 65 countries all these are non-OECD countries excepting Mexico and Korea (who are included in the OECD group in the 1990s). For our analysis the period of our study is chosen to be 1981–2005 in order to get a perfectly balanced panel.

Table 1 Foreign direct investment (FDI) inflows: world total and regional shares, 1970–2007 (selected years)

Regions/years	1970	1980	1990	1995	2000	2003	2005	2006	2007
World total in billion US dollar[a]	13	54	207	341	1398	561	959	1411	1833
Developing economies	29	14	17	34	18	32	33	29	27
Developing economies excluding China	29	14	15	23	15	23	25	24	23
Developing economies: Africa	9	1	1	2	1	3	3	3	3
Developing economies: America (Latin America and Caribbean)	12	12	4	9	7	8	8	7	7
Developing economies: Asia	6	1	11	23	11	21	22	19	17
Developing economies: Eastern Asia	1	2	4	14	8	13	12	9	9
Developing economies: Southern Asia	1	1	0	1	0	1	1	2	2
Developing economies: South-Eastern Asia	3	5	6	8	2	4	4	4	3
Developing economies: Western Asia	1	−6	0	1	0	2	4	5	4
Developing economies: Oceania	1	0	0	0	0	0	0	0	0
High-income developing countries	10	2	9	12	11	13	14	13	12
Middle-income developing countries	13	10	5	6	4	6	8	7	7
Low-income developing countries	7	1	4	15	4	13	11	10	9
Developed economies	71	86	83	65	81	64	64	67	68
Developed economies: Northern America	23	42	27	20	27	11	14	21	19
Developed economies: Asia	1	1	1	0	1	2	1	1	2
Developed economies: Europe	39	40	50	40	52	50	53	42	46
Developed economies: Oceania	8	4	5	4	1	2	−4	2	1

[a]Figures in this row are absolute figures in millions US$ (rounded). Figures in all other rows are percentages of the world total given in this row

Source UNCTAD World Investment Report data available online at http://www.unctad.org

2.1 Granger Causality Tests

First we have examined the causal relationship among the growth rate (PCYG), Trade openness (TRDGDP), and FDI (FDIGDP). For these, we have undertaken VAR (Vector Autoregressive) Granger causality tests. We have considered maximum 10 lags and applied different criteria. These criteria along with the chosen lags (in parentheses) are: sequential modified LR test statistic, LR (9), final prediction error, FPE (9), Akaike information criterion, AIC (9), Schwarz information criterion, SIC (2) and Hannan–Quinn information criterion (2). Evidently there is no consensus; we get two alternative lag values 2 and 9. So, we have applied both 2-order and 9-order Granger causality tests.

From our alternative causality tests it is clear that trade openness (TRDGDP) and the importance of FDI (FDIGDP) are related either by mutual causation or by one-way causality (Table 2). So, in our estimate of the relationship between FDI and growth we have to discard trade openness as an explanatory variable. There is some confusion regarding the direction of causality between FDI and growth: it could be a mutual causal relationship (at lag order 2) or a causality directed from FDI to growth (at lag order 9).

To ascertain the nature of the relationships between FDI and growth we shall now use the panel-version of Pesaran-Shin ARDL methodology.

2.2 Long-Run and Short-Run Relationship Between FDI and Growth: Dynamic Panel Data Analysis

We postulate a long-run relationship between PCYG and FDIGDP:

$$\mathrm{PCYG}_{it} = \beta_i \cdot \mathrm{FDIGDP}_{it} + \eta_{it} \tag{1}$$

where i (=1, 2, … N) represents groups (countries), t (=1, 2, … T) represents periods (years), β_i is the long-run coefficient of FDIGDP, and η_{it} is the error term.

We are interested to know whether there exists a long-run impact of FDIGDP on growth and whether the short-run adjustment dynamics leads to the long-run relationship.

Following Pesaran et al. (1999) our panel data analysis is based on the following error correction representation:

$$\begin{aligned}\Delta\mathrm{PCYG}_{it} = {} & \theta_i\left(\mathrm{PCYG}_{i,t-1} - \beta_i \cdot \mathrm{FDIGDP}_{i,t-1}\right) \\ & + \sum_{j=1}^{p} \lambda_{ij}\Delta\mathrm{PCYG}_{i,t-j} + \sum_{k=0}^{q} \psi_{ik}\Delta\mathrm{FDIGDP}_{i,t-k} + \mu_i + \varepsilon_{it}\end{aligned} \tag{2}$$

Table 2 Growth, trade openness, and net foreign direct investment as % of GDP, FDIGDP, 1981–2005: panel VAR Granger causality tests

VAR Granger causality/block exogeneity wald tests	Dependent variable[a]	Excluded variable[a]	Chi-square	Degree of freedom	Probability
Lag = 2	PCYG	FDIGDP	6.564665	2	0.0375
		TRDGDP	18.26004	2	0.0001
		FDIGDP, TRDGDP	29.40839	4	0.0000
	FDIGDP	PCYG	6.682947	2	0.0354
		TRDGDP	67.01279	2	0.0000
		PCYG, TRDGDP	74.98017	4	0.0000
	TRDGDP	PCYG	7.092644	2	0.0288
		FDIGDP	6.561577	2	0.0376
		PCYG, FDIGDP	12.89827	4	0.0118
Lag = 9	PCYG	FDIGDP	16.97979	9	0.0490
		TRDGDP	29.69403	9	0.0005
		FDIGDP, TRDGDP	51.31855	18	0.0000
	FDIGDP	PCYG	14.31742*	9	0.1115
		TRDGDP	88.08708	9	0.0000
		PCYG, TRDGDP	97.51410	18	0.0000
	TRDGDP	PCYG	8.405573*	9	0.4938
		FDIGDP	12.33178*	9	0.1952
		PCYG, FDIGDP	20.98801*	18	0.2800

[a]*PCYG* Rate of growth real GDP per capita
FDIGDP Net Foreign direct investment as % of GDP
[*]Not significant even at 10 % level. All others are statistically significant

where θ_i is the group-specific error-correcting speed of adjustment term, λ_{ij} and ψ_{ik} are the coefficients of the lagged variables, μ_i is the group-specific effect, and ε_{it} is the disturbances term. The existence of a meaningful long-run relationship with a stable adjustment dynamics requires $\theta_i < 0$.

Using the STATA ado developed by Blackburne et al. (2007) we have estimated the Eqs. (1) and (2). On the basis of Lag Exclusion Wald Test for each variable separately we have determined the lag-structure.[1] Table 3 reports the estimates of the parameters.

[1]We have considered a uniform lag-structure for all the countries as the STATA ado used here does not have this option. It is theoretically possible to consider different lag structures for different countries on the basis of some information criteria.

Table 3 Long-run and short-run effects of foreign direct investment (as % of GDP) on growth and capital formation, 1981–2005: alternative dynamic panel regression estimates

Model[a]	Pooled mean group (PMG)	Mean group (MG)	Dynamic fixed effect (DFE)
A. Dependent variable[a]: growth of GDP per capita, PCYG			
Long-run relationship			
FDIGDP	0.192** (3.37)	0.381 (1.33)	0.238** (4.09)
Short-run relationship			
θ	−0.775** (21.53)	−0.834** (−23.28)	−0.873** (−34.12)
ΔFDIGDP_t	0.291 (1.19)	0.036 (0.12)	−0.037 (−0.67)
$\Delta \text{FDIGDP}_{t-1}$	0.133 (0.81)	−0.014 (−0.07)	0.028 (0.58)
Constant, μ	0.623** (3.43)	0.529 (1.92)	0.667** (4.36)
B. Dependent variable[a]: fixed capital formation (% of GDP), GKFGDP			
Long-run relationship			
FDIGDP	1.062** (10.54)	1.469** (3.26)	0.318** (3.24)
Short-run relationship			
θ	−0.322** (−13.14)	−0.409** (−13.53)	−0.327** (−17.32)
$\Delta \text{GKFGDP}_{t-1}$	0.157** (5.08)	0.176** (5.49)	0.095** (3.78)
ΔFDIGDP_t	0.141 (1.2)	0.128 (1.12)	0.118** (3.28)
$\Delta \text{FDIGDP}_{t-1}$	0.144 (0.86)	0.148 (0.87)	0.059 (1.85)
Constant, μ	5.681** (13.27)	7.657** (13.09)	6.432** (16.69)

*Significant at 5 % level
**Significant at 1 % level
[a]*t*-ratios in parentheses

Following Pesaran et al. (1999) we have used the pooled mean group (PMG) estimator. It allows intercepts, short-run coefficients, and error variances to differ freely across the countries but the long-run coefficients are constrained to be the same; that means, $\beta_i = \beta$ for all i while θ_i may differ from country to country. The PMG estimate supports a positive long-term relationship between FDI (FDIGDP) and growth (PCYG) and there exits a stable adjustment path toward the long-run relationship from the insignificant short-term relationship.

Next we have used the dynamic fixed effect estimators (DFE) where intercepts are allowed to vary across the groups and all other parameters and error variances are constrained to be the same. The DFE estimate tells the same story.

Finally we have used the mean group (MG) estimator where separate equation is estimated for each country and the mean of the estimates is calculated to get a glimpse of the overall picture. This MG procedure finds neither a short-term nor a long-term relationship.

We have also considered the effect of FDI (as % of GDP), FDIGDP on fixed capital formation (as % of GDP), GKFGDP. Now all the three methods of aggregation discussed above, namely, PMG, MG, and DFE tell the same story: a positive impact of FDI on capital formation.

3 Concluding Observation

The decade of 1990s witnessed rapid growth of FDI flows as a part of financial globalization; it declined during 2000–2003 followed by a recovery. This development has a wider support among the economists who are otherwise critical of other aspects of financial globalization such as rising short-term capital flows because of its vulnerability toward financial crisis. Rising flow of FDI is expected to promote growth.

In this perspective this chapter examines the relationship between growth of fixed capital formation and direct foreign investment (as percentage of GDP) in a panel of 65 countries covering a time-period of 1981–2005. It uses the dynamic panel data technique that considers the short-term relationships and the stable adjustment paths leading to the long-term relationships. It aggregates the individual country estimates through the alternative procedures such as mean group, pooled mean group, and dynamic fixed effect. It observes positive long-term effects of FDI on growth in the last two methods and no effect in the first method. That implies that there is no unequivocal support in favor of the proposition that FDI promotes growth. All the methods, however, support the proposition that FDI promotes fixed capital formation. Thus, our study provides a qualified support to Stiglitz (2000): far more relevant (than other types of financial flows) for the long-term success of the economy is foreign direct investment (FDI).

References

Balasubramanyam VN, Salisu M, Sapsford D (1996) Foreign direct investment and growth in EP and IS countries. Econ J 106:92–105

Bende-Nabende A, Ford JL, Santos B, Sen S (2003) The interaction between FDI, output and the spill over variables: co-integration and VAR analyses for APEC, 1969–1999. Appl Econ Lett 10:165–172

Bengoa M, Sanchez-Robles B (2003) Foreign direct investment, economic freedom and growth: new evidence from Latin America. Eur J Polit Econ 19:529–545

Bhagwati J (1998) The capital myth: the difference between trade in widgets and trade in dollars. Foreign Aff 77(May/June):7–12

Blackburne III, Edward F, Frank Mark W (2007) Estimation of non-stationary heterogeneous panels. Stata J 7(2):197–208

Blomstrom M, Lipsey R, Zejan M (1996) Is fixed investment the key to economic growth? Quart J Econ 111(1):269–276

Borensztein E, De Gregorio J, Lee JW (1998) How does foreign direct investment affect economic growth? J Int Econ 45:115–135

Chakraborty C, Basu P (2002) Foreign direct investment and growth in India: a cointegration approach. Appl Econ 34:1061–1073

Choe J II (2003) Do foreign direct investment and gross domestic investment promote economic growth? Rev Dev Econ 7(1):44–57

De Mello LR Jr (1999) Foreign direct investment-led growth: evidence from time series and panel data. Oxford Econ Pap 51:133–151

Durham JB (2004) Absorptive capacity and the effects of foreign direct investment and equity foreign portfolio investment on economic growth. Eur Econ Rev 48:285–306

Egger P, Winner H (2005) Evidence on corruption as an incentive for foreign direct investment. Eur J Polit Econ 21(4):932–952

Gao T (2005) Foreign direct investment and growth under economic integration. J Int Econ 67 (1):157–174

Li X, Liu X (2005) Foreign direct investment and economic growth: an increasingly endogenously relationship. World Dev 33(3):393–407

Lipsey RE (2000) Interpreting developed countries' foreign direct investment. NBER working paper no.7810

Nunnenkamp P (2004) To what extent can foreign direct investment help achieve international development goals? World Econ 27(5):657–678

Pesaran MH, Shin Y, Smith RP (1999) Pooled mean group estimation of dynamic heterogeneous panels. J Am Stat Assoc 94:621–634

Stiglitz J (2000) Capital market liberalization, economic growth, and instability. World Dev 28 (6):1075–1086

Stiglitz J (1994) The role of the state in financial markets. In: Bruno M, Pleskovic B (eds) Annual world bank conference on development economics. World Bank, Washington, D.C., pp 393–410

Foreign Direct Investment and Macroeconomic Indicators in India: A Causality Analysis

Basabi Bhattacharya and Jaydeep Mukherjee

Abstract The objective of the present study is to analyse the pattern of movement of external capital flows to Indian economy in terms of foreign direct investment (FDI) along with other macroeconomic indicators, namely, real GDP, US dollar exchange rate, inflation, T-bill rate, trade openness, and Dow Jones Index value, and their probable impact on the financial and overall performance of the economy. Given that the current study period is from 1997–98 to 2013–14, and the tremendous growth in terms of external capital inflows in India during that time simultaneously with near double digit economic growth, the focus of the chapter is primarily to unravel the linkage between the FDI inflows and the selected macroeconomic fundamentals with the help of quarterly time series data using cointegration and causality analysis.

Keywords Foreign direct investment · Inflation · Trade openness · Growth of real GDP · Toda and yamamoto causality test

JEL Classifications F21 · F23 · C22

1 Introduction

Since the last two decades, emerging economies have been experiencing massive capital inflows which feature as an important dimension of the changing international financial architecture. The two major forms of capital inflows, namely, Foreign direct Investment (FDI) and Foreign Portfolio Investment (FPI) differ in their nature and therefore in their impact on the host economy. This chapter focuses

B. Bhattacharya (✉)
Financial Economics, Department of Economics, Jadavpur University,
Kolkata 700032, India
e-mail: basabi54@gmail.com

J. Mukherjee
Department of Economics, Indian Institute of Foreign Trade, New Delhi, India
e-mail: jaydeepm74@gmail.com; jaydeep@iift.edu

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_18

on the former and intends to analyse its relationship with the macroeconomic indicators in India.

Foreign direct investment refers to foreign investments in the production capacity of the economy. This form of investment is often "preferred over other forms of external finance because they are non-debt creating, non-volatile and their returns depend on the performance of the projects financed by the investors." (Planning Commission of India 2002) Recent Union Budgets of India do emphasize on the enhancement of the investment ambience in India keeping in mind the benefits from FDI through transfer of knowledge, skills and technology and promoting competition in the local product and input market thus raising productivity of the local industry in addition to the other benefits mentioned above. The Union Government keep on trying to uncomplicate the FDI system. Central Government initiatives are revealed towards easing of investment caps and controls in India's high-value industrial sectors—defence, construction and railways, which are now open to global participation. In defence, in particular, the FDI cap is raised from 26 to 49 %, portfolio investment in defence sector is permitted up to 24 % under the automatic route, and 100 % FDI is allowed in defence sector for modern and state of the art technology on case to case basis,. Further, 100 % FDI under automatic route permitted in construction, operation and maintenance in specified rail infrastructure projects, and easing of norms underway for FDI in the construction development sector.

In this regard, it appears important to understand the nature of interlinkage of FDI with India's macroeconomic variables. This is because capital inflows, like FDI, do have multidimensional macroeconomic implications on the host country. In order to reap the due benefit from FDI, it is to be ensured that such capital flows are properly paced and properly sequenced relative to the maturity of the system in which it must be absorbed; this would ensure sustenance of capital flows and systemic stability.

However, there are views emphasizing that the growth impact of FDI is not automatic and is very sector specific. FDI to impart expected benefits is conditional upon good institutions, skilled labour, openness to trade and well-developed financial institutions. Moreover, a study by Gyapong and Karikari (1999) mentions that the FDI leads to higher growth in the countries where FDI is used for the development of the market and the effect is inconclusive when FDI is used for developing the export sector. At the extreme, the opponents view FDI as a new form of 'western imperialism'.

Overtime a voluminous literature has developed in the direction of the determinants of FDI, host country impacts, procedural issues, sectoral and regional distribution within the economy and its contribution to overall economic development. However, a holistic approach in understanding the working of FDI on the host country is required since the FDI works through broad spectrum framework interacting with the macroeconomic variables. The present study addresses this lacuna.

The remaining part of the chapter is organized as follows: A brief review of literature is discussed in Sect. 2. Section 3 explains the methodology and data, Sect. 4

discusses the principal results obtained. The chapter ends with some concluding observations in Sect. 5.

2 The Present State of Art

Literature on FDI has developed mostly along empirical channel though there exists several studies on the theoretical platform as well. The multiple channels through which FDI can work upon the host economy are researched thoroughly. Some major ones are cited herein below.

According to a study by Cho (2003), FDI inflows could bring important benefits to the recipient economies in the form of capital inflows, technology spillovers, human capital formation, international trade integration, enhancement of enterprise development and good governance. FDI may also raise the quality of domestic human capital and improve the know-how and managerial skills of the local firms. Using disaggregated data set on a panel of 14 manufacturing sectors for a sample of developed and developing countries over the period 1992–2004, Cipollina et al. (2012) were able to provide robust evidence on the positive and statistically significant growth effect of FDI in recipient countries. Moreover, they found that this effect is stronger in capital-intensive and technologically advanced sectors. The growth enhancing effect comes primarily from an increase in total factor productivity (TFP) and from factors accumulation. Rodriguez (2000) emphasized upon the fact that smaller the country the more open it should be for rapid economic development. Openness of a country is measured through the extent of international trade.

Mottaleb (2008) used panel data for 60 developing countries in 2003, 2004 and 2005. Sample countries are drawn from the three continents—Asia, Africa and Latin America. His study identifies some features of the host countries such as large domestic market with high GDP growth rate, state of the art infrastructure, business friendly environment and modern communication facilities all of which have encouraged FDI inflow in the developing countries. Similarly, the study by Nonnemberg and Cardoso de Mendonça (2004), using panel data analysis for 38 developing countries (including transition economies) for the period 1975–2000, reveals that GDP and the average rate of growth of GDP in previous years, positively affected the inflows of FDI, being strongly significant. Level of education is found to be highly significant and degree of openness also has positive significance while Inflation and Risk rating of the country showed negative relation with FDI inflows. Among more recent studies, Lipsey and Sjöholm (2011) have exhibited that for Indonesian economy a relatively poor business environment, inefficient government institutions, low levels of education and poor infrastructure all seem to be important explanations for the low inflows of FDI to the country.

Adverse effects of FDI are revealed in Shahbaz and Aamir (2008) study for Pakistan. They have applied ARDL bounds approach for long-run association among macroeconomic variables and established that increased FDI in Pakistan worsens income distribution because it is focused towards capital-intensive industrial and services sectors of urban localities. Karimi and Yusop (2009) using time-series data covering the period 1970–2005 for Malaysia and the Toda-Yamamoto test for causality relationship and the bounds testing (ARDL) approach, found that there is no strong evidence of a bidirectional causality and long-run relationship between FDI and economic growth. FDI could also have negative effects in such areas as market structure and balance of payments and could lead to crowding-out of domestic enterprises. Firm-level empirical studies by Aitken and Harrison (1999) and others, however, have found no positive spillovers running form the foreign firms to the domestic firms. Government policies are therefore needed to enhance benefits and minimize negative effects.

A comparative study of India and China as a destination for FDI flows assumes importance. Keshava (2008) found that China needs to open up the domestic market, better protect intellectual property rights, and improve the performance of state-owned enterprises. While India, which is still far behind China, needs to concentrate more upon infrastructure, power, security considerations, etc., Sharma (2000) draws a comparison between FDI in India and China and tries to find out the probable reasons for a comparatively less impact of FDI on India's economic growth compared to that in China. According to him while China favoured FDI to boost exports rather than promote domestic industries, India's focus was more on its import substitution programme. While China relied more on external sources, India looked more at domestic sources. FDI was seen only as a pipeline for high technology and not as an investment support. So, the policy approach to attract FDI and making best possible use of them is a significant factor working upon the overall impact of FDI.

Jha (2003) emphasized more on the quality of FDI than its quantity. His study shows that FDI in India is not same nationwide but unevenly distributed more towards the developed states rather than the less developed ones like UP and Bihar. The author further identified six major constraints for India's inadequate performance in the area of FDI such as image and attitude, domestic policy, procedures, quality of infrastructure, state government level obstacles and delays in legal processes, which are to be improved to attract more FDI.

The objective of the present study is to complement the existing literature as follows. It undertakes a holistic approach by considering several relevant macroeconomic variables and their interactions with FDI where each chosen variable has its own significance to indicate some economic aspect of the country operating through systemic interactions. It focuses on the linkage between FDI inflows and macroeconomic fundamentals rather than identifying the determinants of FDI inflows as in most of the studies. It investigates the relationship between integrated economic variables by applying the Toda and Yamamoto (1995) causality technique which is superior to traditional Granger (1969) causality test to examine the lead–lag relationships among economic variables.

3 Methodology and Data Sources

3.1 Causality Test

Traditionally Granger (1969) causality is employed to test for the causal relationship between two variables. This test states that, if past values of a variable y significantly contribute to forecast the future value of another variable x then y is said to Granger cause x. Conversely, if past values of x statistically improve the prediction of y, then we can conclude that x Granger causes y. The test is based on the following regressions:

$$y_t = \beta_0 + \sum_{k=1}^{M} \beta_k y_{t-k} + \sum_{l=1}^{N} \alpha_l x_{t-l} + u_t \tag{1}$$

$$x_t = \gamma_0 + \sum_{k=1}^{M} \delta_k y_{t-k} + \sum_{l=1}^{N} \gamma_l x_{t-l} + v_t \tag{2}$$

where y_t and x_t are the two variables, u_t and v_t are mutually uncorrelated error terms, t denotes the time period and 'k' and 'l' are the number of lags. The null hypothesis is $\alpha_l = 0$ for all l's and $\delta_k = 0$ for all k's versus the alternative hypothesis that $\alpha_l \neq 0$ and $\delta_k \neq 0$ for at least some l's and k's. If the coefficient α_l's are statistically significant but δ_k's are not, then x causes y. In the reverse case, y causes x. But if both α_l and δ_k are significant, then causality runs both ways (Bhattacharya and Mukherjee 2008).

The direction of causality depends critically on the number of the lagged terms included. If the chosen lag length is smaller than the true lag length, the omission of relevant lags may cause bias. Conversely, the inclusion of extraneous lags in the equation may cause the estimates to be inefficient. To deal with this problem Hsiao (1981) has developed a systematic autoregressive method for choosing optimal lag length for each variable in an equation. This method combines Granger causality test and Akaike's Final Prediction Error (FPE), defined as the (asymptotic) mean square prediction error.

3.2 Toda and Yamamoto Test

Time series data are often non-stationary and may end up in spurious regression results. Moreover, even when the variables are integrated, the F-test procedure may not be valid, as the test statistics do not have a standard distribution (Gujarati 2006). Toda and Yamamoto (1995) avoid the problems of testing for Granger causality with respect to the power and size properties of unit root and cointegration tests (Zapata and Rambaldi 1997). The procedure requires the estimation of an 'augmented' VAR, even when there is no cointegration, which guarantees the asymptotic distribution of the MWald statistic. The testing procedure is similar to Granger causality, but augmented with extra lags depending on the maximum order of integration of the series under consideration. It is essentially a two-step procedure. First, we need to test for stationarity of the series and also need to find the maximum order of integration ($d_{\max}$). The most popular and widely used test of stationarity is the unit root test, also known as the "augmented" Dickey and Fuller (ADF 1979) test. This test involves estimating the following equation:

$$\Delta y_t = (\phi - 1)y_{t-1} + \sum_{j=1}^{k} \delta_j \Delta y_{t-j} + \varepsilon_t \tag{3}$$

where $\varepsilon_t \sim$ WN (0, σ^2) and then testing for the significance of $(\phi - 1)$.

Sometimes many series contain a drift parameter and a linear trend, and then testing methodology has to be extended in following way. Here we test for the significance of the coefficient $(\phi - 1)$ associated with y_{t-1} in the following regression:

$$\Delta y_t = \beta_0 + \beta_1 t + (\phi - 1)y_{t-1} + \sum_{j=1}^{k} \delta_j \Delta y_{t-j} + \varepsilon_t \tag{4}$$

where, β_0 is the drift parameter.

Once we have checked the stationarity and the level of cointegration then we construct a autoregressive model (VAR) in their levels with a total of $(k + d_{\max})$ lags, where k is the optimal number of lagged terms included which is determined by Hsiao's (1981) FPE method and checked using AIC/SIC criteria. Toda and Yamamoto point out that, for $d = 1$, the lag selection procedure is always valid, at least asymptotically, since $k \geq 1 = d$. If $d = 2$, then the procedure is valid unless $k = 1$. Moreover, according to Toda and Yamamoto, the MWald statistic is valid regardless whether a series is $I(0)$, $I(1)$ or $I(2)$, non-cointegrated or cointegrated of an arbitrary order.

In order to clarify the principle, let us consider the simple example of a bivariate model, with one lag ($k = 1$). That is,

$$x_t = A_0 + A_1 x_{t-1} + e_t \tag{5}$$

or more fully,

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_{10} \\ \alpha_{20} \end{bmatrix} + \begin{bmatrix} \alpha_{11}^{(1)} & \alpha_{12}^{(1)} \\ \alpha_{21}^{(1)} & \alpha_{22}^{(1)} \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \tag{6}$$

where $E(e_t) = E\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = 0$ and $E(e_t e_t') = \sum$

To test that x_2 does not Granger cause x_1, we will test the parameter restriction $\alpha_{12}^{(1)} = 0$. If now we assume that x_{1t} and x_{2t} are $I(1)$, a standard t-test is not valid. Following Dolado and Lutkepohl (1996), we test $\alpha_{12}^{(1)} = 0$ by constructing the usual Wald test based on least squares estimates in the augmented model:

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_{10} \\ \alpha_{20} \end{bmatrix} + \begin{bmatrix} \alpha_{11}^{(1)} & \alpha_{12}^{(1)} \\ \alpha_{21}^{(1)} & \alpha_{22}^{(1)} \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} + \begin{bmatrix} \alpha_{11}^{(2)} & \alpha_{12}^{(2)} \\ \alpha_{21}^{(2)} & \alpha_{22}^{(2)} \end{bmatrix} \begin{bmatrix} x_{1,t-2} \\ x_{2,t-2} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \quad (7)$$

The Wald statistic will be asymptotically distributed as a Chi Square, with degrees of freedom equal to the number of "zero restrictions", irrespective of whether x_{1t} and x_{2t} are $I(0)$, $I(1)$ or $I(2)$, non-cointegrated or cointegrated of an arbitrary order.

3.3 Data Measurement

The current study deals with the data during the period 1997–98 to 2013–14. Any time series study should preferably have as frequent data as possible, especially in the case when nature of data is very dynamic. But given the fact that weekly or monthly data for many macroeconomic indicators are not available, quarterly data have been used for the study, with data for GDP (base year prices of 1993–94) and WPI (at base year 1993–94 = 100) based on common base period. All variables are measured in local currency (INR) and are expressed in natural logarithms. The data source is Reserve Bank of India's (RBI) Handbook of Statistics on Indian Economy.

3.4 Macroeconomic Indicators

The indicators selected to understand the macroeconomic determinants of Foreign Direct Investment (FDI) flow in India were gross domestic product (GDP), US dollar Exchange rate in Rupees (ERATE), inflation (WPI), index of industrial production (IIP), trade openness (TOPEN), 364 days T-bill Rate (INTEREST), and Dow Jones Index (DJINDEX) values.

- Value of Foreign Direct Investment (FDI) inflow is taken as a dependent variable. With the help time series data, the impact of various other parameters on FDI inflow will be checked.

- Gross domestic product (GDP) is taken as proxy for economic development of India considering that higher growth rate of the real GDP will attract more FDI inflow for India. This is taken with the consideration that better resources capability can attract more FDI inflow.
- Dollar exchange rate (ERATE) exchange rate is taken as proxy to check how volatility of exchange rate of Indian Rupees with respect to currency of developed countries can affect the flow of FDI to India.
- Inflation (WPI) acts as a proxy for the level of economic stability, considering that one of the classic symptoms of loss of fiscal or monetary control is unbridled inflation. WPI is taken with the consideration that investors prefer to invest in more stable economies.
- Trade openness (TOPEN) is taken as a proxy for the type of relation India maintains with foreign nations in terms of the capital. TOPEN is taken with the consideration that a greater degree of trade openness demonstrate India's acceptance for foreign capital and foreign goods and services. TOPEN is defined as (Imports + Exports)/GDP.
- 364 day T-bill rate (INTEREST) is taken as proxy for base (risk free) return.
- Dow Jones Index value (DJINDEX) is taken as proxy for developed market portfolio return; with the consideration if the returns in developed financial markets are more than the return achieved through direct investment in a developing country like India then it would make FDI in India less attractive.

4 Empirical Results

The results of Augmented Dickey Fuller (ADF 1979) unit root test are depicted in Tables 1 and 2. The results suggest that the logarithm of FDI and all the macroeconomic indicators are stationary, except (log of) real GDP. However, the first difference of real GDP is stationary, i.e., integrated of order 1. Thus, the variables in the model are not integrated of same order and hence the traditional F-test in Granger causality may not be reliable in inferring leads and lags among such variables, with different orders of integration (Toda and Phillips 1993).

4.1 *Toda-Yamamoto Causality Test*

Since the variables are integrated of different orders, in our study, the causality has been tested through Toda-Yamamoto approach. As observed in Tables 1 and 2, that the maximum order of integration ($d_{\max}$) equals 1; we next determine the number of lagged terms (k) to be included using AIC/SIC rule and find it to be two. Finally, we construct a VAR in levels, similar to that depicted in Eq. (7) with a total of ($k + d_{\max}$) equalling three lags.

Table 1 Results for unit root tests in levels

Variables	ADF	
	Constant no trend	Constant with trend
LFDI	−7.43*	−7.51*
LGDP	−1.02	−1.85
LERATE	−5.24*	−6.03*
LWPI	−5.91*	−6.33*
LTOPEN	−2.83***	−2.91***
LINTEREST	−8.56*	−8.89*
LDJINDEX	−7.32*	−7.27*

Note Asterisk (***) and (*) denote statistically significant at 10 and 1 % level, respectively

Table 2 Results for unit root tests in first difference

Variables	ADF	
	Constant no trend	Constant with trend
LGDP	−11.65*	−11.83*

Note Asterisk (*) denotes statistically significant at 1 % level

$$\begin{bmatrix} \mathrm{FDI}_t \\ \mathrm{GDP}_t \\ \mathrm{CRATE}_t \\ \mathrm{ERATE}_t \\ \mathrm{WPI}_t \\ \mathrm{TOPEN}_t \\ \mathrm{INTEREST}_t \\ \mathrm{DJINDEX}_t \end{bmatrix} = B_0 + B_1 \begin{bmatrix} \mathrm{FDI}_{t-1} \\ \mathrm{GDP}_{t-1} \\ \mathrm{CRATE}_{t-1} \\ \mathrm{ERATE}_{t-1} \\ \mathrm{WPI}_{t-1} \\ \mathrm{TOPEN}_{t-1} \\ \mathrm{INTEREST}_{t-1} \\ \mathrm{DJINDEX}_{t-1} \end{bmatrix} + B_2 \begin{bmatrix} \mathrm{FDI}_{t-2} \\ \mathrm{GDP}_{t-2} \\ \mathrm{CRATE}_{t-2} \\ \mathrm{ERATE}_{t-2} \\ \mathrm{WPI}_{t-2} \\ \mathrm{TOPEN}_{t-2} \\ \mathrm{INTEREST}_{t-2} \\ \mathrm{DJINDEX}_{t21} \end{bmatrix} + B_3 \begin{bmatrix} \mathrm{FDI}_{t-3} \\ \mathrm{GDP}_{t-3} \\ \mathrm{CRATE}_{t-3} \\ \mathrm{ERATE}_{t-3} \\ \mathrm{WPI}_{t-3} \\ \mathrm{TOPEN}_{t-3} \\ \mathrm{INTEREST}_{t-3} \\ \mathrm{DJINDEX}_{t-3} \end{bmatrix} + e_t$$

where B_0 is the intercept vector and e_t is the vector of error terms.

The results of the Toda-Yamato tests of Granger causality are in Table 3. The causality test results suggest bidirectional causality between trade openness and FDI at 1 % significance, and unidirectional causality from inflation to FDI at 10 % significance. The inflation rate is a key indicator of monetary and fiscal policy of a country. Changes in inflation rates of the domestic or foreign country are anticipated to make adjustments the net returns and optimal investment decisions of the multinational enterprises. A low and controlled inflation environment often encourages FDI inflow. It may be noted that the inflation rate in India (based on WPI) was moderate at below 5 % during most of the period under study.[1] The very first year of the Eleventh Plan (2007–12) has seen a resurgence of inflation in the last quarter mainly due to the steep rise in international prices of oil and food grains. Controlling the perils of price rise has been one of the major challenges for the

[1]Average rate of inflation (based on WPI) was 4.9 % during the Ninth Plan (1997–98 to 2001–02) and 4.8 % during the Tenth Plan (up to January, 2006) [Source: Economic Survey of India 2006–07: http://indiabudget.nic.in/es2006-07/esmain.htm].

Table 3 Results of causality test due to Toda-Yamamoto (1995) Procedure

Null hypothesis	MWALD statistic	*p*-values
Foreign Direct Investment (FDI) versus Gross Domestic Product (GDP)		
GDP does not Granger Cause FDI	1.35	0.15
FDI does not Granger Cause GDP	0.68	0.49
Foreign Direct Investment (FDI) versus Dollar Exchange Rate (ERATE)		
ERATE does not Granger Cause FDI	1.83	0.27
FDI does not Granger Cause ERATE	1.82	0.38
Foreign Direct Investment (FDI) versus Inflation (WPI)		
WPI does not Granger Cause FDI	2.73***	0.06
FDI does not Granger Cause WPI	0.52	0.18
Foreign Direct Investment (FDI) versus Trade Openness (TOPEN)		
TOPEN does not Granger Cause FDI	4.36*	0.00
FDI does not Granger Cause TOPEN	4.82*	0.00
Foreign Direct Investment (FDI) versus 364-days T-bill Rate (INTEREST)		
INTEREST does not Granger Cause FDI	0.61	0.72
FDI does not Granger Cause INTEREST	0.54	0.57
Foreign Direct Investment (FDI) versus Dow Jones Index (DJINDEX)		
DJINDEX does not Granger Cause FDI	0.43	0.81
FDI does not Granger Cause DJINDEX	0.28	0.76

Note Asterisk (***), (**) and (*) denote statistically significant at 10, 5 and 1 % level, respectively

government over the last decade, with the inflation averaging 8.6 % over the past 3 years. The average WPI inflation reached a 4-year low of around 6 % in 2013–14, providing a relief. The downward trend has been caused primarily by reduction in inflation in non-food manufactured products due to fall in global commodity prices, which came down to a 4-year low of 2.9 % in 2013–14.

5 Conclusion and Policy Implications

The present study is an endeavour to investigate the interlinkage of FDI and the impact on the macroeconomic variables in India by exploring causal relationship, if any, between the former and the latter through econometric analysis on quarterly time series data for the period 1997–08 to 2013–14. While the Granger two-step procedure test has been applied to test for cointegration among FDI and macroeconomic variables, Granger non-causality proposed by Toda and Yamamoto (1995) has been used for causality test.

Our results indicate that FDI and the selected macroeconomic indicators are cointegrated and thus there exists a long-run stable relationship between them. The short-run causality test result suggests bidirectional causality between trade openness and FDI, and unidirectional causality from inflation to FDI inflows. The bidirectional causality between trade openness and the FDI inflow easily understood by India's gradual opening up to capital inflows and can be established as a positive feedback loop to increase FDI inflow to India. Causality from inflation to FDI inflows is interpreted to reflect that moderate inflation with the possibility of growth can attract FDI inflow.

Some general observations might appear relevant. First of all, the quality of FDI is equally, if not more, important than its quantity. FDI becomes attractive for its own sake when it makes a net contribution to the economic development of the host country. Policy should target FDI which '…triggers technology spillovers, assists human capital formation, contributes to international trade integration and particularly exports, helps create a more competitive business environment, enhances enterprise development, increases total factor productivity and, more generally, improves the efficiency of resource use' (OECD 2002), rather than any FDI, per se. The better the quality of FDI more beneficial would be the impact on the macroeconomic field of the host country. This would, of course, have to be associated with the performance of the sectors or industries where FDI flows in and linkages these industries have with other development channels of the economy. Second, regional balance in FDI flow within an economy is also important. Among the major states in India Maharastra, Karnataka, Tamil Nadu and Gujarat are the dominant recipients; West Bengal, Andhra Pradesh, Madhya Pradesh and Orissa have received considerably lower share with Bihar and UP having even less. In order to have the desired macroeconomic impact through FDI, need-based prioritization of FDI across sectors and regions within a country is required. Ensuring quality FDI and appropriate orientation of FDI flow within an economy would gradually promote more effective interlinkage between the macroeconomic variables and the FDI. If the recent policy initiatives for FDI promotion in India are enriched with the above considerations, more powerful long-run relationships between FDI and macroeconomic variables may be hoped for.

References

Aitken BJ, Harrison Ann E (1999) Do domestic firms benefit from direct foreign investment? Evidence from Venezuela. Am Econ Rev 89(3):605–618

Bhattacharya B, Mukherjee J (2008) Indian equity market since liberalization: efficiency, volatility and structural break. World View, Kolkata

Cipollina M, Giovannetti G, Pietrovito F, Pozzol Alberto F (2012) FDI and growth: what cross-country industry data say. World Econ 35(11):1599–1629

Cho J-W (2003) 'Foreign direct investment: determinants, trends in flows and promotion policies'. Chapter V of investment promotion and enterprise development bulletin for Asia and the pacific, 1, Economic and Social Commission for Asia and The Pacific, UN

Dickey DA, Fuller WA (1979) Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc 74(366):427–431

Dolado JJ, Lutkepohl H (1996) Making wald test work for cointegrated VAR systems. Econometric Rev 15(4):369–386

Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37(3):424–438

Gujarati DN (2006) Essential of econometrics, 3rd Edn. McGraw Hill
Gyapong AO, Karikari JA (1999) Direct foreign investment strategies and economic performance in Ghana and Ivory Coast. J Econ Dev 24(1):133–147
Hsiao C (1981) Autoregressive modelling and money-income causality detection. J Monetary Econ 7(1):85–106
Jha R(2003) "Recent trend in FDI flows and prospects for India", Unpublished in Social science research network electronic paper collection. http://ssrn.com/abstract=431927
Karimi MS, Yusop Z (2009) "FDI and economic growth in Malaysia", MPRA Paper No. 14999. http://mpra.ub.uni-muenchen.de/14999/
Keshava SR (2008) "The effect of FDI on India and Chinese economy; a comparative analysis", Unpublished in Social Science Research Network (SSRN). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1089964. Accessed 2 Feb 2008
Lipsey Robert, Sjöholm Fredrik (2011) Foreign direct investment and growth in East Asia: lessons for Indonesia. Bull Indonesian Econ Stud 47(1):35–63
Mottaleb AK (2008) 'Determinants of foreign direct investment and its impact on economic growth in developing countries', Unpublished in MPRA Paper No. 9457
Nonnemberg MB, Cardoso de Mendonça MJ (2004) "Determinants of foreign direct investment in developing countries". www.anpec.org.br/encontro2004/artigos/A04A061.pdf
OECD (2002) "Foreign direct investment for development: maximizing benefits minimizing costs", Report of the steering group on foreign direct investment: foreign investment India. OECD, Paris. http://www.oecd.org/dataoecd/47/51/1959815.pdf
Planning Commission of India (2002) "Report of the steering group on foreign direct investment: foreign investment India". Planning Commission, Government of India, New Delhi. http://planningcommission.nic.in/aboutus/committee/strgrp/stgp_fdi.pdf. Accessed 10 June 2005
Rodriguez AC (2000) Macroeconomic policy lessons from LDC's. J Appl Economics 3(1): 169–216
Shahbaz M, Aamir N (2008) Direct foreign investment and income distribution: a case study for Pakistan. Int Res J Financ Econ 21(2008):7–18
Sharma K (2000) "Export growth in India: has FDI played A Role?". Economic Growth Center, Yale University, Center Discussion Paper No. 816
Toda HY, Phillips PCB (1993) Vector autoregressions and causality. Econometrica 61(6): 1367–1393
Toda HY, Yamamoto T (1995) Statistical inference in vector autoregressions with possibly integrated processes. J Econometrics 66(1–2):225–250
Zapata HO, Rambaldi AN (1997) Monte carlo evidence on cointegration and causation. Oxford Bull Econ Stat 59(2):285–298

Part V
Issues Relating to Globalization, Financial Markets and Financial Instruments

Exploratory Study of Select Commodity and Equity Indices Around the Meltdown of 2008

Diganta Mukherjee and Arnab Mallick

Abstract Many believe that the world of investment is becoming complicated and more volatile with every passing day. What makes it even more interesting is the complexity of the market particularly for those select instruments which are traded on electronic exchanges, be it equity, commodity or currency, to name a few. Each investment market has its own dynamics and their inter-linkages are even more exciting.

1 Lay of the Land

Many believe that the world of investment is becoming complicated and more volatile with every passing day. What makes it even more interesting is the complexity of the market particularly for those select instruments which are traded on electronic exchanges, be it equity, commodity or currency, to name a few. Each investment market has its own dynamics and their inter-linkages are even more exciting. A huge amount of research has been conducted in areas related to equity and currency market, but over the last decade, tremendous amount of interest has been generated around commodity market as well. Although, historically one of the oldest markets known to mankind, exchange (electronic) traded commodity market is rather a relatively recent phenomenon in several key countries, for example, exchange traded commodity market has been operating in India since 1875.

But amidst all development and penetration in the financial world, the incident that has shaken the whole investment world has been the global financial meltdown in 2008 triggered by the subprime mortgage crisis. It has been nearly 7 years since the impact could be felt and understood and plenty of water has flown since then. The world has witnessed unprecedented monetary policy boosters from economic

D. Mukherjee (✉)
Indian Statistical Institute, Kolkata, India
e-mail: diganta@isical.ac.in

A. Mallick
Indian Institute of Foreign Trade, New Delhi, India

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_19

giants including the United States of America with China and Japan joining the group rather recently. Very few would like to debate that the meltdown and the series of action thereafter to tackle the global meltdown has had a major impact on the investment market. This article is an attempt to touch upon certain softer aspects of the shift in select commodity and equity market spectrum around that period. In this article, an attempt has been made to provide certain critical observation of select markets which have been severely impacted from the meltdown, purely from an investment point of view. Although, we have tried to provide a holistic picture to provide a sense of developments in the market, there has been effort to link it to different kind of markets (in terms of maturity of the market). Focusing mostly on base metals market, this work has made reference to several other exchange traded commodity baskets like gold, as well key leading equity indices across the work to give a feel of how the recent developments has been influencing these key platforms of investments.

Meanwhile, before getting deep into the topic, it is worth mentioning that the very basic of commodity market has one very unique feature which gives an added layer of complexity to this market. Trends of prices for exchange traded commodities are not entirely determined by equilibrium situation from physical demand and supply configuration but largely influenced by flow of financial funds into the select spectrum purely driven by investment motives. To elaborate it in somewhat more details although, market forces including demand and supply have a huge impact in determining the spot price of commodities but in today's electronically integrated financial world, in most cases traders take cues from the futures price of commodity contract (in cases, where they exist) and in real world it tends to influence spot market sentiments. Professional investors do not take their positions on commodity-based contracts (on the futures exchanges) only on the fundamental factors like demand, supply, inventory in physical market, but on host of other factors like currency, economic environment, alternative investment opportunities at particular time point, technical indicators arising out of a historic price movement, risk appetite, amongst several others. In reality, often signals from futures market end up having a great influence on physical traders of the commodities and thereby impacts the spot prices. Thus, in reality one often sees a both way causal relationship between spot prices and futures prices. The fact that price of exchange traded commodity prices are often influenced by host of factors in addition to fundamentally linked factors like physical demand and supply, can be well felt in the recent rise in volatility of certain commodity basket; for example gold—an extremely critical commodity as historically it is also treated as a hedge against uncertainty and therefore extremely popular amongst a large section of investor community. The recent correction in gold price and sharp rise in volatility is difficult to explain on the back of general theory of market economics only. This very interesting phenomenon has been gaining popularity amongst researcher for long. It has been vividly explored and discussed by Mayer (2009) financial investors regard commodities as an asset class (comparable to equities, etc.) and do not necessarily trade on the basis of fundamental supply and demand relationships in specific commodity markets.

In particular, topic related to exchange traded commodity markets are gaining importance amongst researcher and financial professionals as the growing volatility particularly post the recent financial meltdown is pretty significant. In particular, the commodities segment including energy, base metal and precious metals is in a critical phase at the moment as most of the commodities are trading at near year lows now after sharp decline and thereafter a sharper recovery post the financial meltdown in 2008, clearly suggesting that the basket might have lost investor's interest in competition to its close counterpart, i.e. the equity market during the respective time frame, or perhaps, market fundamentals surrounding commodity basket complex are clearly not in favour at the moment.

But, at this junction the million dollar question remains whether the current market situation is a result of normal market function or is it a result of fund flows into commodity complex for investment or trading requirements?

The initial sharp upturn in price immediately post the heavy doses of quantitative easing from economic giants in response to initial meltdown to the financial crash in 2007–08 was not unexpected. There was huge amount of quantitative easing which was pumped into the market primarily by the US followed by a series of other measures by key sovereign entities as everyone was trying to frame the best possible solution to combat the on-setting downturn. The financial meltdown had caused massive destruction in market capitalization and consequently, jittery investors were waiting on every possible instance to make money through their investments, result of which was clearly visible in the equity and commodity market as it rallied sharply once it got its first chance. Moving pretty much in line with other asset classes, exchange traded commodities also showed strength.

With passage of time, things started getting tricky and are getting seriously difficult at the moment as prices are collapsing fast, and with an impact which is far reaching, and also one of the basic reasons why this work focuses more on commodities to see how has key and select commodities moved in recent past post the financial meltdown.

It is worth noting that leading corporations are often associated in business pertaining to several key commodities, but whose prices are largely determined by the commodity exchanges. Thus, these prices therefore are subject to major influence of financial fund flows mostly from financial communities having little or no interest in the physical commodity per se, for consumption or value addition, as has been mentioned in the past.

As a result, market participants with a commercial interest in physical commodities (i.e. producers and consumers) are facing increased uncertainty about the reliability of signals emanating from the commodity exchanges. This makes the argument even more contemporary today as business environment is taking sharp turns as downward cycle in commodity price are forcing global leading miners like BHP Billiton, Rio Tinto, and Vale amongst several others to significantly shift their operation strategies which have resulted into sharp reduction in capital and exploration expenditure for next few years. Needless to say, the impact is far more severe with strong linkage effect and impacting global economic recovery, particularly for countries whose economic activities are primarily rotating around commodity markets, like Australia, Indonesia, Brazil, Chile, amongst several others.

Therefore this is perhaps an appropriate time to once again look into the dynamics of the exchange traded commodity market, particularly with a focus to understand as how fund flows between alternative sources of investment are creating havoc in certain pockets in the commodities market post the financial crisis of 2007. As more and more investors are getting exposed to different kinds of investment avenues including commodities, this topic is gaining increasing relevance. It may either be a direct involvement of large financial institutions or indirect involvement of retail investors through different products with an exposure into the commodity complex.

Along with key commodity like gold (which has a unique positioning amongst investors) this exploratory study will lay much focus on base metal segment due to some key reasons. A commodity basket like base metals has a huge final consumption demand as well because of its huge industrial usage. Once again as mentioned in the past, prices of these commodities in the physical market are directly linked with price discovered on select leading commodity exchanges, which (futures price) is largely a result of speculative fund flow from large financial institutions. Consequently, many believe that there is an apparent increase in disconnect between physical market fundamentals (actually demand and supply of the commodity) with the prevailing prices and surely the emerging prevailing volatility. Needless to say, it is emerging as a bigger threat for stability of those fundamental commodities, and hence once again, its relevance in the globally integrated commodity market is getting bigger with every passing day. At the same time, its relevance for nearly all countries (which are either importers and/or exporters) of this basket of commodity on futures exchange is also gaining a lot of focus. It is also worth noting that currently, there are just a very few essential commodities which are not traded on commodity exchange. But, it will not be wise to assume that the situation will remain like this forever. Hence, an understanding of the dynamics of exchange traded commodities is extremely relevant in the context of international business.

It is not only important for countries which are exposed to these commodity price fluctuation risks but also for different entities including companies involved with these commodities in some way or the other and surely the global population at large. This exploratory study also aims at finding these interesting movements across critical investments basket post the great financial shock and tries to pose certain key question which could be looked into greater details at consequent studies, primarily focusing around exchange traded commodity market.

2 Background Literature

The scale at which commodity space is attracting investor interest is relatively rather a recent phenomenon and several recent researches in this regards provide evidence of its growing dominance. Vrugt et al. (2004) have shown in their work that for a long time, commodities were deemed inappropriate investments because of their perceived risky character. The disappointing performance and future prospects of traditional asset classes and the availability of data and commodity indices have rapidly changed this situation.

Commodities as an investment destination have consequently gained a lot of interest in academic research. There has been plenty of evidence that also supports the incentives of several economic agencies getting increasingly interested in commodities complex.

Kazemi et al. (2009), Georgiev (2001) have explored in detail the benefits of commodity investment. Vrugt et al. (2004) has showed that investors can profit from tactical asset allocation with commodities in real-time and argued that the timing strategy delivers superior investment returns, both in an economical and a statistical sense. Conover (Georgiev 2001) has explored the question when commodities need to be added to someone's investment portfolio.

A very interesting work in the commodities spectrum has been done by Jacquier et al. (1999) in which they provided evidence that commodities perform well when the general financial market climate is negative. Interestingly, given the current global environment of quite a bit of uncertainty, better to say, confusion, commodity spectrum has of late been an underperforming asset class.

Another unique aspect in commodities research was provided by Bodie (1983), Froot (1995), Gorton and Rouwenhorst (2004) wherein it was explained that commodities appeared to serve as a possible hedge against inflation, which makes them even more attractive for entities with fixed liabilities in real terms, like for instance pension funds. Nijman and Swinkels (2003) show that commodity investments are beneficial to pension funds within a mean variance framework. As widely accepted, historically gold has always been considered to a natural hedge against inflation. However, recent trend shows that the yellow metal has been consistently losing price on the courses with occasional flips. This price decline is surely difficult to explain purely on basis of demand-supply changes over last few years.

Understanding of price dynamics of exchange traded commodities are no longer a matter of interest of academic research only. As mentioned in the past, it is perhaps of utmost importance as never before considering the impact that it has been showing to have on major economies which are largely driven by commodities cycle and given the interdependence of different economies in today's trading environment, impact of commodity price is much higher than before. Gruss (2014) has explored the matter in detail for LAC and concluded that the end of the commodity price boom will entail a significant drag on growth for the average commodity exporter of LAC. Needless to suggest that this concretely has a larger implication on policy making which could therefore needs to be framed to minimise the impact.

In any discussion pertaining to global commodity market mostly metals, the name that always secures topmost slot is China, along with the United States of America. The tremendous consumption appetite of China has clearly emerged as one of the most critical determinants of physical market fundamentals. Also, as time has passed, the Chinese commodity exchange, name the Shanghai Futures Exchange (SHFE) has gained massive attention, surely in the metal space. As an obvious expectation, there has also been tremendous volume of research around this zone and researchers have found unique insights. Roache (2012) in his work to understand

China's impact on commodity markets with a focus on the spillover of aggregate activity and commodity-specific demand shocks has shown that shocks to aggregate activity in China have a significant and persistent short-run impact on the price of oil and some base metals. Interestingly the study also finds that China's impact on world commodity markets is rising but remains smaller than that of the United States. This is mainly due to the dynamics of real activity growth shocks in the U.S, which tend to be more persistent and have larger effects on the rest of the world. It would be interesting to explore this dynamics in terms of a real business cycle model augmented with expectations regarding the relative importance of countries. In fact, if the participating economies believe that USA is more important in terms of leading trade patterns, then in equilibrium this would be a self-fulfilling expectation.

Therefore understanding of the pattern, particularly the change that market may have undergone post the recent financial meltdown is of gaining interest to many, finally leading to evaluate the pattern of linkage of a matured market with relatively newer but extremely crucial emerging markets.

There has been some work in this regard in the equity market, but very few in the commodity space with a focus on post meltdown period. Crucial work has been done by Verma (2005) wherein findings suggested that the magnitude and duration of an upturn in the US market are fully reflected in equity markets of Latin America and that the impact is significantly different from that of a downturn.[1] This is no longer a surprise in today's globalised world with volatility cross clustering being often prevalent.[2] Meanwhile, with the financial turmoil hitting the global market really hard, the market has seen change in reaction to crucial parameters including risk appetite of investors, movements of fund flow, sensitivity towards currency movement, positioning in the derivatives market, reaction to monetary policy changes, amongst many.

Particularly for base metal what is also very interesting is the inter-linkage between select key exchanges on which those commodities are traded. Investors and traders closely follow prices in different markets to take appropriate positions in exchanges of interest. Historically and traditionally, base metals have been long traded on London Metal Exchange (LME) and COMEX with LME setting the lead as benchmark prices. Off late with emergence of Chinese dominance on the commodity complex, the latest addition to this elite list of exchanges has been Shanghai Futures Exchange (SHFE). As very expected the recent economic turmoil has also resulted in jerk in the way market participants behaves in the market.

All these three exchanges along with India's Multi Commodity Exchange (MCX) interestingly reflect four significant strata of global economy in terms of 'economic power centre'. With the US market is fighting hard to maintain its dominance in global commodity market as China has been moving ahead fast to

[1]This interestingly verifies the reference point dependent utility functions as proposed by Kahneman and Tversky.

[2]This phenomenon has been verified by econometricians using multidimensional asymmetric GARCH models. For a discussion see Campbell et al. (1996).

possibly replace the US as the most critical component in deciding equilibrium situation in commodity complex, mainly metals. The UK market is unique in the context of base metals, because of the existence of the LME but otherwise a weak economy at the moment, and finally India, a market which is in everyone's radar as a fastest emerging market. Similar movement shift is also felt in terms of exchange traded commodity market. There is a growing consensus among market participants that SHFE is emerging fast and may be slowly replacing LME which has traditionally been the most crucial benchmark for the base metal complex. If that happens, it will have immense significance in the scheme of things.

With growing debate on possible shifts in centre of financial and economic power centres of the world, which is an obvious result of several years of macroeconomic development in respective countries, there is also a change in the way market participant is giving weightage to prices movement on different exchange. In case of base metals, it is now a growing competition between SHFE and LME as who would take the centre position amongst investor's mind.

Some work has been done around these areas of market inter-linkages. Long and Lei (2008) has worked on dynamic relationship among China's metal futures, spot price and London's futures price. Researchers in the past have explored the linkage between markets with a special focus of China. Zhao (2002) has studied in detail the degree of association between Chinese commodity futures market with international futures market. In one of the very crucial studies done by Garrigues et al. (2014) it has been indicated that it is Comex, followed by the SHFE, not the LME which plays the most important role in copper price discovery. Consequently, relatively newer exchanges have emerged including Multi Commodity Exchanges (MCX) and NCDEX in India. What has been really interesting is the linkage between these markets and the nature of their relationship.

3 Exploratory Data Analysis of Select Price Trends

It is worth noting that the fate of commodity market on back of global financial meltdown was similar to that of the equity market at large. But, its impact on common people's investment kitty (directly) was not felt as it has been the case of equity as exposure of common people to commodity basket as an investment destination that has been relatively low as compared to equity. But things are changing at a rapid pace and it is emerging as a major investment destination in the current financial world. Apart from the mere financial aspect, the implication of this emerging trend is significant. The emerging volatility and price appreciation followed by sharp correction is having a major effect on the macroeconomic fundamentals of most economies.

To begin with, investigating the gold price trend over past few years reveals some very interesting insights. Traditionally treated as a sticky investment option, gold price has rarely seen such a correction as it has been witnessing now (see Fig. 1). In particularly, gold has seen one of the sharpest corrections over past 2 years. Interestingly, during the phase of the correction, there was no major shift in

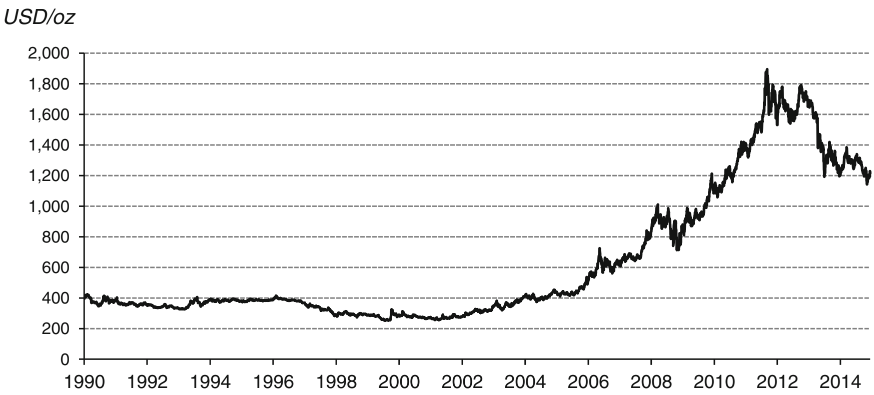


Fig. 1 Gold spot price in US dollars, London PM fix. *Source* Reuters Datastream, LBMA, World Gold Council

the fundamental forces in the physical market, e.g. demand and supply of gold market. It was more a function of financial fund flow that caused this sharp price movement. A large section of professional investors believe that this kind of price movement are more to do with fund flows as compared to factors to do with physical market realities, like demand and supply (see Table 1).

Silver, another popular member in precious metal basket traded on popular commodity exchanges, has witnessed a much sharper rise of around 339 % from $10.24/oz in November 2008 to $45.05/oz in April 2011, thereafter correcting sharply by 62 % to $17.28/oz in October 2014. Although these commodities do not have a direct impact on the lives of common people, there are plenty of commodities (with much larger impact of common people's daily life) which have experienced nearly a 100 % rise in price in as short as 2 years period. However, there is very little evidence of physical market fundamentals supporting the same.

In reality, post beginning of the recent economic turmoil in 2007 and subsequent price crash in 2008, the financial market has witnessed a huge inflow of additional funds as result of major 'quantitative easing' policy adopted by several leading economies, with the United States of America leading the list. A portion of the fund also got routed toward the commodity basket of the world resulting into a sharp pull back in price. This sudden rise in commodity price perhaps ended up giving a deceiving signal to a segment of corporation involved in those sectors leading massive investment pouring in some key segment. In effect this injection of funds created a commodity market bubble. Eventually, there seems to have developed a perennial market distortion that has been created and prices of commodities came crashing down and have been hovering at few year lows, leaving those corporation who had already invested quite a bit caught in the wrong foot (see Fig. 2).

Coming back to the fundamental factors once again, the huge interest of investor community has also perhaps led to sharp increase in volatility in several commodity baskets as fundamental scenario does not change so often so as to have its ramification on price so frequently. It was also coupled with the decline in momentum of

Table 1 World Gold Market: Demand and Supply

	2012	2013	Q4′12	QV′13	Q2′13	Q3′13	Q413	Q1′14	Q2′14	Q3′14	Q3′14 versus Q3′13 % chg
Supply											
Mine production	2,869.7	3,054 0	746.4	690.6	738.5	802.3	822.5	723.9	756.7	812.0	1
Net producer hedging	−39.7	−40.0	−31.7	−10.6	−15.1	−7.5	−6.8	7.6	55.0	−15.0	–
Total mine supply	2,829.9	3,014.0	714.6	680.1	723.4	794.9	815.7	731.6	811.6	797.0	0
Recycled gold	1,633.7	1,242.1	390.7	359.4	253.4	333.7	295.5	306.6	250.1	250.5	−25
Total supply	4,463.7	4,256.1	1,105.3	1,039.5	976.8	1,128.6	1,111.2	1,038.2	1,061.7	1,047.5	−7
Demand											
Fabrication											
Jewellery[1]	2,0073	2,369.6	510.6	545.1	729.7	567.8	527.0	565.2	508.7	541.9	−5
Technology	415.3	408.2	98.0	103.5	103.5	103.1	98.1	98.1	100.5	97.9	−5
Sub-total above fabrication	2,422.6	2,777.8	608.6	648.6	833.3	670.9	625.0	663.4	609.1	639.8	−5
Total bar and coin demand	1,343.4	1,773.3	380.8	462.9	632.8	312.3	365.3	282.6	266.0	245.6	−21
ETFs and similar products[2]	279.1	−880.0	88.1	−176.5	−402.2	−120.2	−181.0	−2.6	−39.9	−41.3	–
Central bank net purchases[3]	544.1	409.3	150.4	130.8	92.1	101.5	85.0	124.3	117.8	92.8	−9
Gold demand	4,589.2	4,080.5	1,227.9	1,065.8	1,155.9	964.4	894.4	1,067.6	953.0	937.0	−3
OTC investment and stock flows[4]	−125.5	175.6	−122.7	−26.3	−179.1	164.2	216.8	−29.5	108.8	110.5	−33
Total demand	4,463.7	4,256.1	1,105.3	1,039.5	976.8	1,128.6	1,111.2	1,038.2	1,061.7	1,047.5	−7
London PM fix (US$/oz)	1,6690	1,411.2	1,721.8	1,631.8	1,414.8	1,326.3	1,276.2	1,293.1	1,288.4	1,281.9	−3

Source World Gold council

[1] Jewellery fabrication: Fabrication is the first transformation of gold bullion into a semi-finished or finished product. Jewellery consumption is equal to fabrication plus/minus jewellery imports/exports plus/minus stocking/de-stocking by distributors and manufacturers

[2] Exchange Traded Funds and similar products including: Gold Bullion Securities (London), Gold Bullion Securities (Australia), SPDR® Gold Shares (formerly streetTRACKS Gold Shares), NewGold Gold Debentures, iShares Comex Gold Trust,ZKB Gold ETF, GOLDIST, ETF Securities Physical Gold, ETF Securities (Tokyo), ETF Securities (NYSE), XETRAGOLD, Julius Baer Physical Gold, Central Fund of Canada and Central Gold Trust, Swiss Gold, iShares Gold Bullion Fund (formerly Claymore Gold Bullion ETF), Sprott Physical Gold Trust, ETF Securities Glitter, Mitsubishi Physical Gold ETF and iShares Gold CH

[3] Excluding any delta hedging of central bank options

[4] Partly a statistical residual, this data is largely reflective of demand in the opaque over-the-counter (OTC)

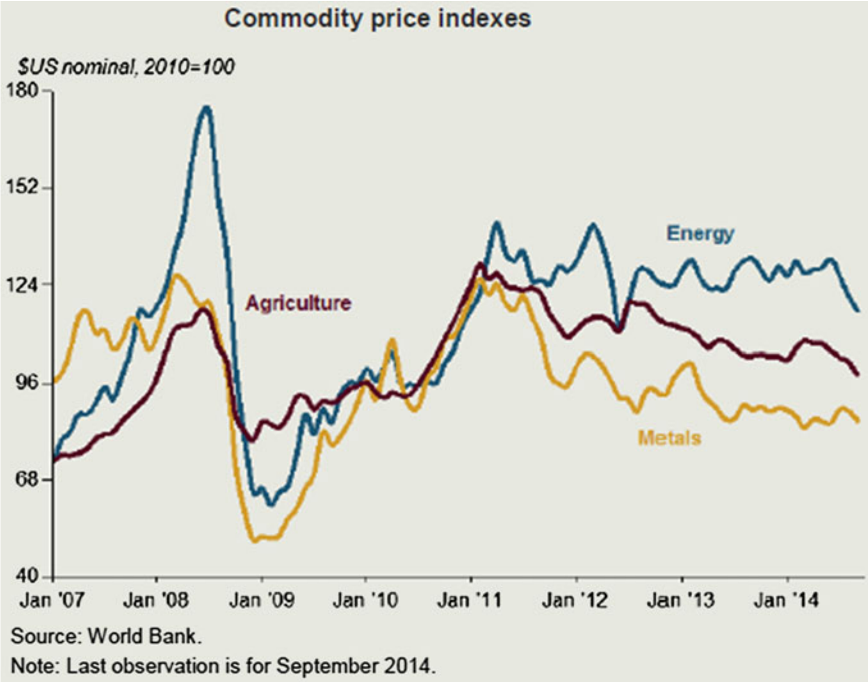


Fig. 2 World Bank Commodity Price Indexes

pickup of demand from China and eventually decline in certain segment. This deteriorated the investment interest pulling down price further.

To focus on metal segment, the commodity market outlook of the World Bank, October 2014 confirms that the World Bank metals price index reached a high of 126 in February 2011 (2010 = 100), up 164 % since its December 2008 low (see Fig. 3).

This upturn together with the increase prior to the financial crisis, generated large new investments and a strong supply response with large amount of the additional metal supply finding its way to meet Chinese demand, whose consumption share of world refined metals reached 47 % at the end of 2013, up from 45 % in the previous year (and up from 5 % two decades ago). Given its share in the world consumption basket, China has been one of the most critical countries of interest to all. Many analysts started arguing that the upward phase of the commodity "super-cycle" that started in the early 2000s has run its course.

The decline in prices was, however, modestly halted in 2014Q3, with the World Bank metals price index rising 2.6 % (q/q). Base metals drove the increase in prices (up 5.3 %, q/q) while iron ore prices experienced a steep drop (down 12 % q/q). Iron ore prices are down for the third quarter in a row, reflecting expansion of low cost producers, particularly Australia by leading giant BHP Billiton and Rio Tinto. The rise in prices of base metals reflects expectations of tightening supply

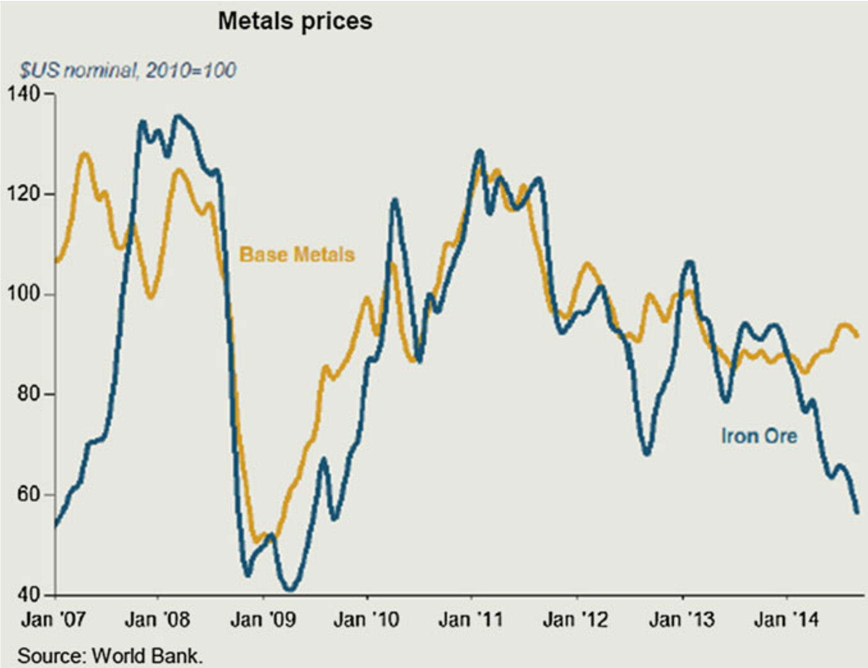


Fig. 3 World Bank Metal Price Index

conditions, which have since dissipated and reversed course by the end of the quarter, says the World Bank Report confirming once again the fact that speculation and expectation have been having a strong influence on price formation in these crucial counters. On the demand side, Chinese imports have weakened as growth of imports of aluminium, zinc, copper and iron ore has slowed to zero or turned negative in three months to August, 2014 (see Fig. 4).

Graphs showing overall performance of commodity including precious metal, energy basket and also some leading equity indices over last few years centering around the financial meltdown (see Figs. 5 and 6).

Graph showing the daily percentage returns of the key metals and S&P500 (see Fig. 7).

It is easy to observe that there is a serious spike in volatility after the 2006 meltdown period in all the graphs above except for Zinc. The disturbance shows persistence for a reasonably long period (couple of years). Thus, it is a moot question whether there is evidence of volatility clustering. We explore this in terms of two simple sets of calculations as shown below. The correlations between the daily returns of base metals are quite substantial and their relation with S&P500 is also not ignorable. So, there is sufficient evidence of concordance in directional change (see Table 2).

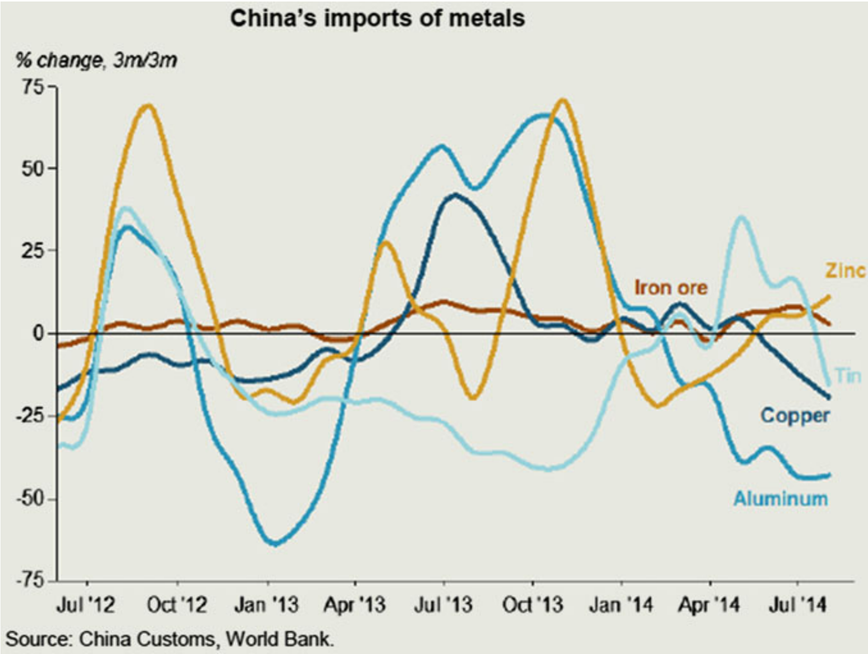


Fig. 4 China's import of metals: Quarterly growth trend

A more interesting observation emerges when we look at the second set of correlations between the squares of the daily returns to gauge the concordance of volatility. It is seen that copper, aluminium and nickel are tightly bound with each other in terms of volatility clustering, whereas zinc is more autonomous. The index is also reasonably closely bound with copper and nickel (see Table 3).

The above calculations indicate that the persistence of volatility has been self-reinforcing in a group of metals which also implied the long term persistence observed in the graphs above.

4 A Few Researchable Ideas

It is quite well known that both spot and future prices are complexly influenced by a host of factors like currency, inflation, M3, interest rate, inventories, availability of arbitrage window, changes in fundamental factors including expected demand-supply mismatch amongst several others. Amongst several complex aspects involved in commodity market, another extremely important factor that has significant impact is the currency exchange rates. Since fund from across the world gets invested in commodity complex in select exchanges in different countries, cross country currency exchange often emerges as a key determinant of net return

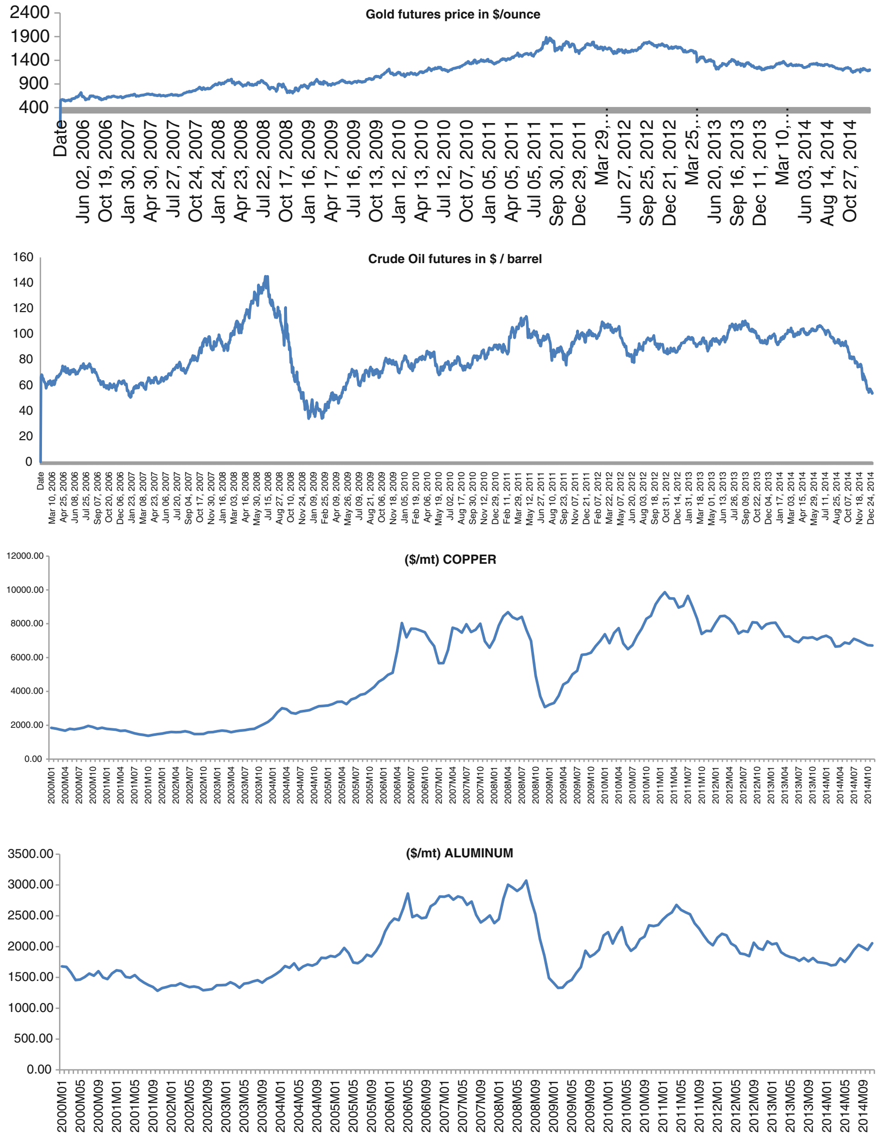


Fig. 5 Overall performance of some leading exchange traded commodities of the world. *Source* Investing.com; World bank

one seeks to get from an invested position. Needless to say, the dynamics of the global commodity market is changing very fast. Sensitivity in price, be it in the spot market or futures, to different factors is not the same anymore. A significant amount of work has been devoted towards understanding these price movements. In a nutshell, any further looking at ‘change in sensitivity of commodities price towards

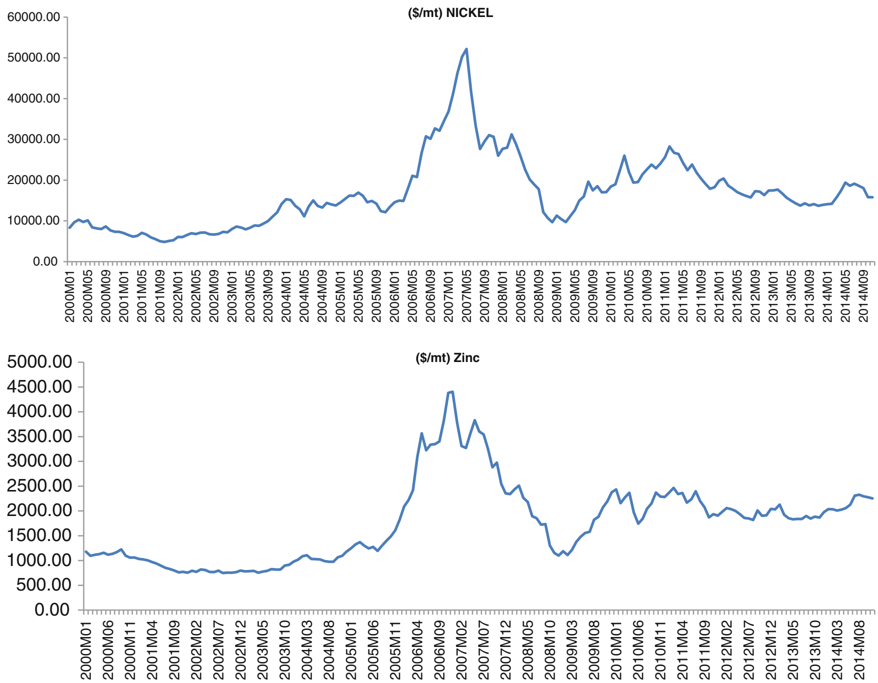


Fig. 5 (continued)

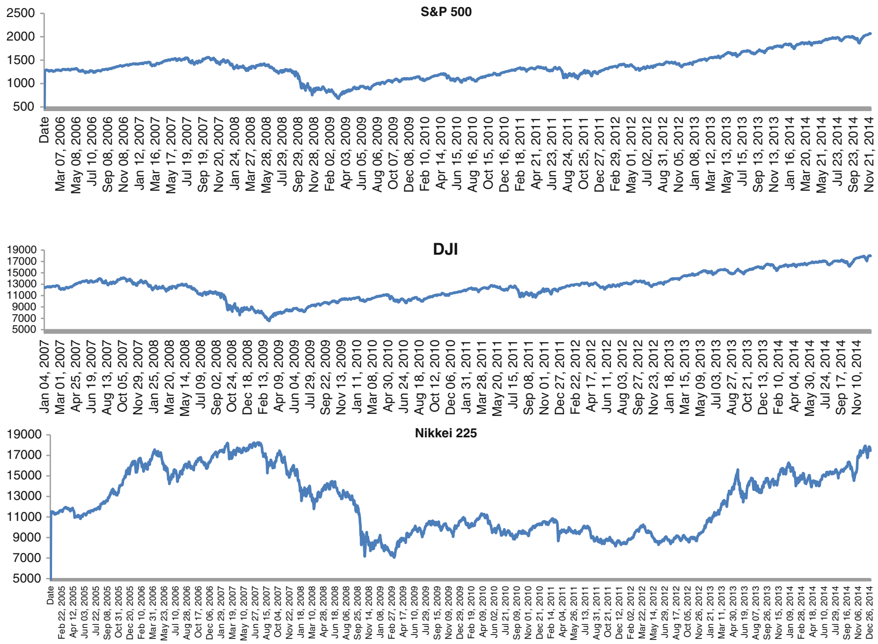


Fig. 6 Overall performance of some leading equity indices of the world. *Source* Investing.com

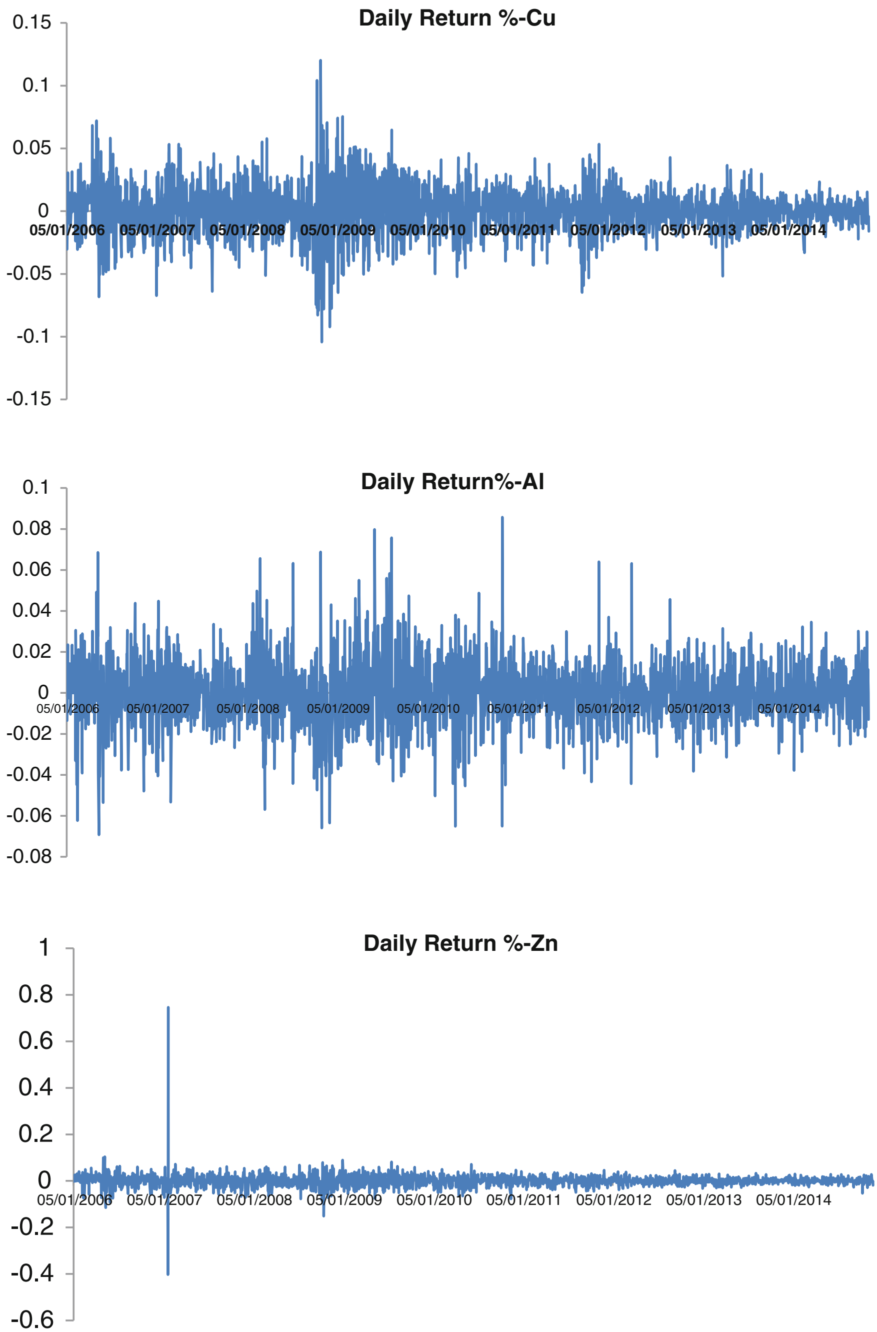
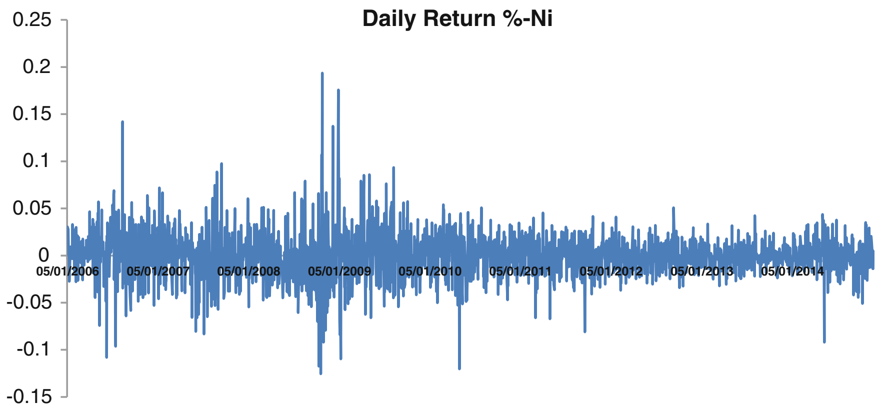


Fig. 7 Graph showing the daily percentage returns of the key metals and S&P500



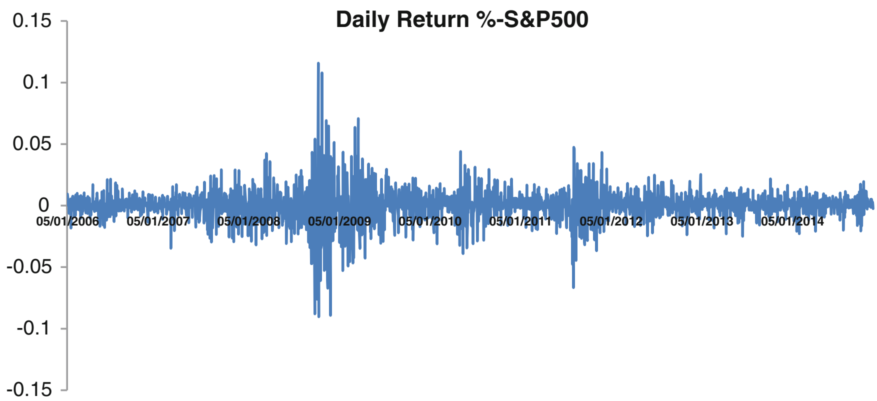


Fig. 7 (continued)

Table 2 Correlation between daily returns

	Cu	Al	Zn	Ni	S&P500
Cu	1.00				
Al	0.67	1.00			
Zn	0.58	0.51	1.00		
Ni	0.61	0.55	0.49	1.00	
S&P500	0.16	0.13	0.12	0.14	1.00

crucial macroeconomic and metal specific factors as well as overseas markets post the financial meltdown' could be a very interest area to look into.

Another most crucial aspect of the financial world is impact of information. In today's world of information technology and communication revolution, special

Table 3 Correlation between squared daily returns

	Cu	Al	Zn	Ni	S&P500
Cu	1.00				
Al	0.45	1.00			
Zn	0.04	0.04	1.00		
Ni	0.52	0.32	0.08	1.00	
S&P500	0.13	0.04	0.01	0.11	1.00

emphasis will also be given to explore to what extent shocks in form of 'market news/event study' (temporary or permanent-long term) impact the market sentiment and influence price. Consequent to the recent financial meltdown, the investment community has become very shaky and the elasticity with which it is responding to any macroeconomic/geopolitical/fundamental developments has been much sharper. In this section, this crucial aspect of the market (which also determines to a large extent the short term volatility) will be studied. Interestingly, Odhan (1998) has examined markets in which price-taking traders, a strategic-trading insider, and risk-averse market-makers are overconfident. The paper shows markets also under react to abstract, statistical, and highly relevant information, and they overreact to salient, anecdotal, and less relevant information. Choi et al. (2014) have explored by assuming the block exogeneity of US macroeconomic variables with respect to the international non-fuel commodity prices, and how exogenous macroeconomic shocks affect those commodity prices. Further studies of the trend in response profile of market participants of exchange traded commodity market to news/market information, post financial meltdown, is something which would add significant value to understanding the underlying dynamics of this market.

In a globally integrated environment with near free flow of information, there is a requirement to analyse how emerging but extremely strategic markets like India respond to global signals. Bose (2008) has explored the Indian commodity market in detail. The work has attempted to bring forth the nature of information flow between futures and spot prices in the market for commodity derivatives in India, taking into consideration the history of commodity derivatives globally, and the importance of and problems associated with commodity markets particularly in less mature economies. However, it has not looked into inter-linkage amongst global market and linkage of Indian market with other major exchanges. Jagadharini and Putran (2014) has studied issues that one of the major commodities has been facing. Further studies in this area to explore the extent of lagged impact that major exchanges individually and/or collectively have on the Indian market is surely worth exploring. Measurement of this predictability will have a significant impact on the investment appetite and fund flow into the Indian market as there will then be an opportunity of risk free profit. "Predictability of the Indian markets and possibility of systematic profit booking through understanding of lagged impact other matured market" would be a massive breakthrough finding in the Indian context.

References

Bodie Z (1983) Commodity futures as a hedge against inflation. J Portfolio Manage 12–17

Bose S (2008) Commodity futures market in India: a study of trends in the notional multi-commodity indices. Money Finance

Campbell JY, Lo AW, Mckinlay AC (1996) Econometrics of financial markets. Princeton University Press

Choi K, Hammoudeh S, Kim WJ (2014) The effects of U.S. macroeconomic shocks on the international commodity prices

Chow G, Jacquier E, Kritzman M, Lowry K (1999) Optimal portfolios in good times and bad. Financ Anal 65–73

Froot KA (1995) Hedging portfolios with real assets. J Portfolio Manage 60–77

Garrigues IFF, Gilbert CL, Yan J (2014) Copper price discovery on comex, the LME and the SHFE, 2001–2013. Working paper, Departamento de Economía de la Empresa

Georgiev G (2001) Benefits of commodity investing. CISDM working paper, March 2001, University of Massachusetts

Gorton G, Geert Rouwenhorst K (2004) Facts and fantasies about commodity futures. NBER working paper, nr. 10595

Gruss B (2014) After the boom–commodity prices and economic growth in Latin America and the Caribbean. IMF Working paper WP/14/154 by

Jagadharini TS, Putran R (2014) Issues before a national multi commodity exchange

Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. Econometrica 47:263–291

Kazemi H, Schneeweis T, Spurgin R (2009) The benefits of commodity investment. AIA Research Report, Alternative Investment Analytics Llc, Aug 2009

Long R, Lei W (2008) Research on the dynamic relationship among china's metal futures, spot price and London's futures price. Int J Bus Manage 3(5)

Mayer J (2009) The growing interdependence between financial and commodity markets. United nation discussion paper

Nijman TE, Swinkels LAP (2003) Strategic and tactical allocation to commodities for retirement savings schemes, CentER discussion paper, nr. 20

Odean T (1998) Volume, volatility, price, and profit when all traders are above average. J Financ 53(6):1887–1934

Roache SK (2012) China's impact on world commodity markets. Working paper WP/12/115

Verma R (2005) Do emerging equity markets respond symmetrically to US market upturns and downturns? Evidence from Latin America, Int J Bus Econ 4(3):93–208

Vrugt E, Bauer R, Molenaar R, Steenkamp T (2004) Dynamic commodity timing strategies. Limburg Institute of Financial Economics

Zhao J (2002) Cointegration analysis of commodity futures price in Dalian, China and Chicago. Dongbei University of Finance & Economics, Dalian, China

An Empirical Investigation of Volatility Clustering, Volatility Spillover and Persistence from USA to Two Emerging Economies India and China

Ayanangshu Sarkar and Malabika Roy

Abstract The issues of volatility and risk in recent times have gained importance for financial practitioners, market participants, regulators and researchers. Volatility is the most basic statistical risk measure instrument. This chapter empirically investigates the pattern of volatility in the Indian and Chinese stock markets during 2006–2011 with reference to its time varying nature, presence of certain characteristics such as volatility clustering and existence of 'spillover effect' in the domestic and the US stock markets. This chapter will also try to estimate the persistence of shock in terms of half-life in each sub-period of study. It contributes to the body of knowledge by providing a holistic outlook to the subject of stock market volatility in India and provides evidence on its main features with the help of econometric techniques employing GARCH models. A comparative analysis is made with the Chinese stock market taking Shanghai Composite Index (SCI).

Keywords Volatility clustering · Volatility spillover · Leverage effect · GARCH (1,1)

1 Introduction

Volatility is the most basic statistical risk measure for financial economics. Stock return volatility is the standard deviation of daily stock returns around the mean value and the stock market volatility is the return volatility of the aggregate market portfolio. This chapter takes into consideration the stock market volatility of two

A. Sarkar (✉)
Department of Business Administration,
Pailan College of Management & Technology, Kolkata, India
e-mail: ayanangshu2@rediffmail.com

M. Roy
Department of Economics, Jadavpur University, Kolkata, India
e-mail: malabikar@gmail.com

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_20

emerging economies India and China to make a comparative analysis of the pattern of clustering effect, leverage effect and spillover effect.

Daily stock prices have been converted to daily returns. The study uses the logarithmic difference of prices of two successive periods for the calculation of rate of return. The logarithmic difference is symmetric between up and down movements and is expressed in percentage terms for ease of comparability with the straightforward assumption of a percentage change.

Let I_t be the closing level of SENSEX on date t and I_{t-1} be the same for its previous business day, i.e. omitting intervening weekend or stock exchange holidays, then the one-day return on the market portfolio is calculated as

$$r_t = l_n(I_t/I_{t-1})\,100$$

where, *ln(z)* is the natural logarithm of '*z*'.

One of the key assumptions of the ordinary regression model is homoscedasticity, which says that the errors have the same variance throughout the sample. If the error variance is not constant, the data are said to be heteroscedastic. Findings of heteroscedasticity in stock returns are well documented by Mandelbrot (1963), Fama (1965) and Bollerslev (1986). These studies have found that stock return data is typically characterized by: serial correlation indicating that successive returns are not independent, serial correlation in the squares of returns resulting in distinct periods of high volatility and relative stability, i.e. volatility clustering and negative asymmetry in the distribution of returns questioning the assumption of an underlying normal distribution. Leptokurtosis in the distribution of returns with too many values near the mean and in the tails of the distribution when compared with the normal distribution.

In econometric literature, volatility clustering is modelled as an ARCH process. Engle (1982), in his seminal work on inflation in the UK, first introduced the idea of ARCH effect. Later on, Bollerslev (1986) generalized this type of model and introduced the GARCH model. The ARCH and the GARCH models assume conditional heteroscedasticity with homoscedastic error variance. That is, the changes in variance are a function of the realizations of preceding errors and these changes represent temporary and random departures from a constant unconditional variance. The advantage of GARCH model is that it captures the tendency in financial data for volatility clustering. It, therefore, enables us to make the connection between information and volatility—since any change in information arrival to the market will change the volatility in it. Thus, unless information remains constant, which is hardly the case, volatility must be time varying even on a daily basis.

Conversions of daily stock prices into daily returns have been first processed by using Microsoft Excel. Subsequently, econometric analysis package EViews has been used to test the return and volatility data for various statistical properties and to estimate ARCH/GARCH class of models.

The chapter is organized as follows: in Sect. 2, we report the results of relevant diagnostic tests. In Sect. 3, we report the results of GARCH model, which gives an idea of persistence of shock. In Sect. 4, report the EGARCH results. Section 5 concludes.

2 Diagnostic Tests

Descriptive statistics on SENSEX and SCI returns are summarized in Table 1. For both SENSEX and SCI, the Skewness statistic for daily returns is found to be different from zero indicating that the return distribution is not symmetric (0.124538) in case of SENSEX and (−0.427937) in case of SCI. Furthermore, the relatively large kurtosis (SENSEX—8.755812; SCI—5.311320) suggests that the underlying data is leptokurtic or heavily tailed and sharply peaked about the mean when compared with the normal distribution. The Jarque–Bera statistic calculated to test the null hypothesis of normality rejects the normality assumption. The results confirm that daily stock returns are leptokurtic and skewed (Table 2).

The autocorrelation (AC) and partial autocorrelation (PAC) were computed. The highly significant Ljung-Box-Pierce Q-statistic, as shown in Table 3 and Table 4, confirms the presence of first order correlation in the return series and negates random walk behaviour. Diagnostics for AR and MA models confirm AR (1) structure of the mean equation for both SENSEX and SCI. The above findings indicate the possible presence of ARCH effect which is confirmed by the computed value of Lagrange Multiplier (LM). This finding shows the clustering effect in daily returns, i.e. large shocks to the error process are followed by large ones and small shocks by small ones of either direction. The existence of a leptokurtic distribution,

Table 1 Descriptive statistics of the daily return series

	RTINDIA	RTCHINA	RTUSA
Mean	0.018245	0.028100	0.002697
Median	0.052504	0.067676	0.030390
Maximum	6.944362	3.923526	4.381650
Minimum	−5.039746	−4.019897	−3.776172
Std. dev.	0.852045	0.885067	0.648456
Skewness	0.124538	−0.427937	−0.118923
Kurtosis	8.755812	5.311330	9.416482
Jarque-Bera	1710.740	313.1036	2124.950
Probability	0.000000	0.000000	0.000000
Sum	22.56913	34.75976	3.336267
Sum sq. dev.	897.3128	968.2133	519.7320
Observations	1237	1237	1237

Table 2 Correlations of returns of SENSEX, SCI. DJ

	RTINDIA	RTCHINA	RTUSA
RTINDIA	1.000000	0.023681	0.060653
RTCHINA	0.023681	1.000000	−0.013522
RTUSA	0.060653	−0.013522	1.000000

Table 3 Autocorrelation and partial autocorrelation of SENSEX

Autocorrelation and partial autocorrelation of SENSEX				
	AC	PAC	Q-Stat	Prob.
1	0.115	0.115	16.320	0.000
2	0.153	0.142	45.405	0.000
3	0.135	0.108	68.157	0.000
4	0.206	0.170	121.04	0.000
5	0.136	0.080	144.19	0.000
6	0.116	0.046	160.99	0.000
7	0.160	0.093	192.69	0.000
8	0.093	0.010	203.43	0.000
9	0.136	0.060	226.37	0.000
10	0.196	0.132	274.30	0.000

Table 4 Autocorrelation and partial autocorrelation of SCI

	AC	PAC	Q-Stat	Prob.
1	0.009	0.009	0.0394	0.843
2	0.065	0.065	2.1490	0.341
3	0.203	0.203	22.851	0.000
4	0.036	0.032	23.493	0.000
5	0.001	−0.026	23.494	0.000
6	0.079	0.035	26.616	0.000
7	0.124	0.119	34.431	0.000
8	0.017	0.017	34.575	0.000
9	0.030	−0.007	35.046	0.000
10	0.183	0.140	52.181	0.000

volatility clustering and a changing conditional variance means that our next logical step in the modelling exercise should be to express the conditional volatility as an ARCH or GARCH process with the mean return as an AR (1) process.

2.1 Unit Root Tests

Financial markets react nervously to shocks or crisis periods. Statistically speaking, it means that the conditional variance for the given past is not constant over time and the process X_t is conditionally heteroskedastic. The stationarity or otherwise of a series can strongly influence its behaviour and properties, e.g. persistence of shocks will be infinite for nonstationary series. We have used both Dickey and Fuller test and Phillips–Perron test to check the stationarity of SENSEX and SCI before proceeding to GARCH modelling. The tests are similar to ADF tests, but they incorporate an automatic correction to the DF procedure to allow for auto-correlated residuals.

Table 5 Unit root test

	Augmented Dickey–Fuller	Phillips–Perron
RTINDIA (SENSEX)	−32.71278 (0.0000)	−32.73662 (0.0000)
RTCHINA (SCI)	−35.47560 (0.0000)	−35.49863 (0.0000)

The results of both the tests confirm that the series are stationary. Table 5 presents the results of these tests.

2.2 *Chow Test*

To study the nature of stock market returns and volatility for two emerging economies: China and India, we divided the period of study into different sub-periods of stock market upswings and stock market downswings depending on pre and post recessionary period. These periods are

Period-1: 4th March 2006–14th December 2007
Period-2: 17th December 2007–1st April 2009
Period-3: 2nd April 2009–31st March 2011

These periods are determined by plotting the level data and observing the stock market upswings and downswings depending on pre- and post-recessionary period. Then, we applied Chow's Break Point test to check whether there were any structural breaks on these dates.

Table 6 Chow's breakpoint result for SENSEX

F-statistics	927.9433	Prob. F(4,1216)	0.0000
Log likelihood ratio	1716.370	Prob. Chi-Square(4)	0.0000
Wald statistics	3711.773	Prob. Chi-Square(4)	0.0000

Table 7 Chow's breakpoint result for SCI

F-statistics	328.9185	Prob. F(4,1216)	0.0000
Log likelihood ratio	896.0980	Prob. Chi-Square(4)	0.0000
Wald statistics	1315.650	Prob. Chi-Square(4)	0.0000

To run a Chow's breakpoint test, we have to estimate a regression equation on the data on which we are looking for the break point. For this purpose, we separately run a least square regression equation taking the level data of SENSEX(India) and SCI(China) as dependent variable and taking DowJones (USA), as independent variable. India = $C(1) + C(2)$ * USA, CHINA = $C(1) + C(2)$ * USA are the two simple regression equations are estimated then Chow's break point test is applied on them.

Results of Chow break points test furnished in Tables 6 and 7 corresponding to two break points as decided on 14th December, 2007 and 1st April, 2009.

Null Hypothesis: No breaks at specified breakpoints.

We found that for both SENSEX and SCI Chow's breakpoint test rejected the null hypothesis that there exist no breaks at specified level. So in the rest of the analysis, we have presented a comparative analysis for three of these sub-periods to understand time changing pattern of volatility.

3 GARCH Model: Results

3.1 GARCH Model to Establish Persistence of Shock and to Calculate Half-Life of Shock

Bollerslev (1986) developed the GARCH model.

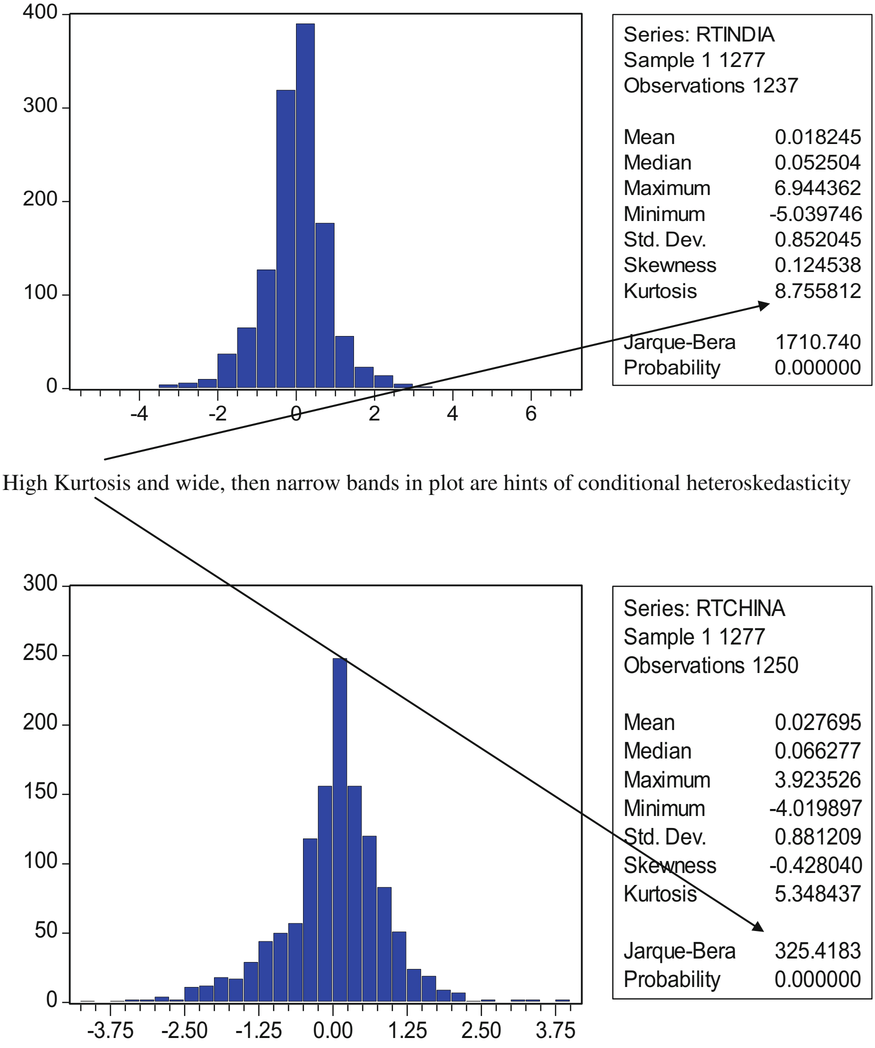
GARCH (1,1):

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

The variance (h_t) is a function of an intercept (ω), a shock from the prior period (α) and the variance from last period (β).

High Kurtosis and wide, then narrow bands in plot are hints of conditional heteroskedasticity.

Diagnostic Test for GARCH and GARCH Estimates



The parameter estimates of the GARCH (1,1) models in Table 8 are all statistically significant. The estimates of β_1 are always markedly greater than those of α_1 and the sum $\beta_1 + \alpha_1$ is very close to but smaller than unity. The fact that $\beta_1 + \alpha_1$ is close to unity, however, is useful for purposes of forecasting conditional variances. The sum, however, is rather close to one, indicates a long persistence of shocks in volatility. The closer the value of $\beta_1 + \alpha_1$ to 1 the longer will be the persistence of shock. We observe that $\beta_1 + \alpha_1$ is closest to 1 in the final sub-period. Thus in the

Table 8 GARCH estimates of SENSEX and SCI (for detail calculation refer Appendix 1)

	Period-1: 4th March 2006–14th December 2007		Period-2: 17th 2007–1st April 2009		Period-3: 2nd April–31st March 2011	
	SENSEX	SCI	SENSEX	SCI	SENSEX	SCI
A	0.098926 (0.0007)	0.128874 (0.0004)	−0.058058 (0.3254)	−0.162750 (0.0000)	0.050780 (0.0229)	0.029746 (0.34498)
ω	0.026160 (0.0002)	0.013697 (0.0021)	0.0751856 (0.0571)	0.035248 (0.0000)	0.005644 (0.0321)	0.023166 (0.0384)
α_1	0.180966 (0.0000)	0.061899 (0.0000)	0.131614 (0.0072)	−0.040575 (0.0000)	0.118561 (0.0000)	0.042569 (0.0044)
β_1	0.775294 (0.0000)	0.924676 (0.0000)	0.820384 (0.0000)	1.018870 (0.0000)	0.875519 (0.0000)	0.905456 (0.0000)
$\alpha_1 + \beta_1$	0.956260	0.986575	0.951998	0.978295	0.994080	0.948025

third period when the SENSEX was recovering from its lowest value, persistence of volatility increased significantly. The reasons may be that the investors became more cautious and increased their risk premium attached to stocks that they cannot consider the shock to be transitory and short term in nature any more. The result in case of China is not altogether similar to India. In case of SCI, we find the value of $\beta_1 + \alpha_1$ is lowest in third period and the value was 0.9458 indicating that shock persistence had reduced in the third sub period compared to first two.

These results are further consolidated by calculating the half-life of shock in all the periods in the next section. Throughout the analysis, the value of $\beta_1 + \alpha_1$ is less than unity, which indicates no violation of the stability condition. Poterba and Summers (1986) have argued that for a long period asset like stocks, persistence of shocks is needed to be able to explain the time varying risk premium. The reason for such an argument is that if shocks are only transitory in nature, i.e. has only short-term effect, investors will not make any changes in their discounting factor while obtaining the present discounted value of the stock and hence its price.

3.2 Half-Life Calculation

Lamoureux and Lastrapes (1990) have proposed a half-life period of a shock to the variance. Half-life period is that period in which the shock diminishes to half of its original size. The half-life for GARCH (1,1) process is $1 - [\log e2/\log e(\alpha 1 + \beta 1)]$.

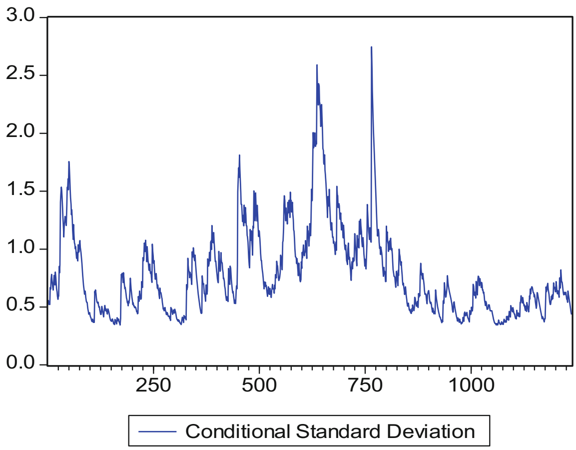
Table 9 shows that the half-life of shock in SENSEX in three sub-periods are 17, 16 and 118 days, respectively, indicating persistence of shock has increased drastically during the post recessionary period of recovery. Thus, the effect of a shock to the volatility process of daily return takes about maximum 117 days to diminish by half its original impact in the case of SENSEX during the final recovery period. This result is consistent with our earlier findings of GARCH estimate proving significantly high

Table 9 Half-life of shock calculation for the indices in three sub-periods

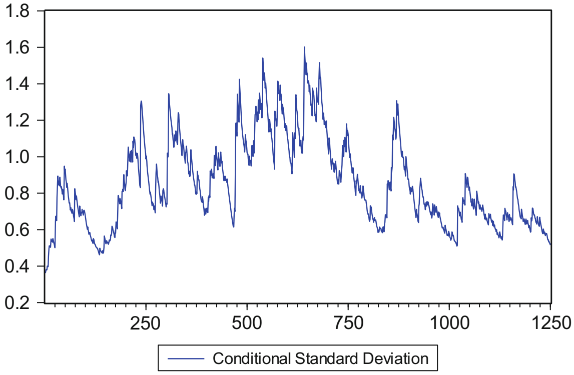
	Period-1: 4th March 2006–14th December 2007		Period-2: 17th 2007–1st April 2009		Period-3: 2nd April–31st March 2011	
	SENSEX	SCI	SENSEX	SCI	SENSEX	SCI
α_1	0.180966	0.061899	0.131614	−0.040575	0.118561	0.042569
β_1	0.775294	0.924676	0.820384	1.018870	0.875519	0.905456
Half-life	16.50	52.28	15.09	32.59	117.74	13.99

persistence parameter for the third sub-period. The result is somehow opposite in case of SCI China. The half-lives for three periods are 53, 33 and 14 days, respectively, indicating that it has successfully hedged the risk of shock persistence in the path of recovery from the world recession. The reason may be the listed industries of SCI have reduced their integration with other stock markets of the world. Also, China has always attracted more FDI than India. This is because China's policies for foreign investors are more liberal than India. Moreover, the Chinese economy is growing faster and infrastructure is better. Although strict protection policies remain in place in China, in selected sectors such as automobiles, India's restrictive labour laws and limits affecting foreign shares in ownership restrain foreign investment in general. And in particular, India's inadequate infrastructure development makes it very difficult for multinational companies to ship products in and out of the country, and even within the country. It appears that any bad or good news does have a significant and long lasting impact on the volatility of the stock prices.

In the following section, we have plotted the conditional standard deviation graph of SENSEX and SCI. Conditional sd graph shows brief periods of high volatility for both India and China. But if we look at the scale, the magnitude of volatility found to be much higher in case of India compared to China. Where the volatility of SENSEX crosses 2.5 % for more than few times, in case of China it didn't cross the 1.6 % mark.

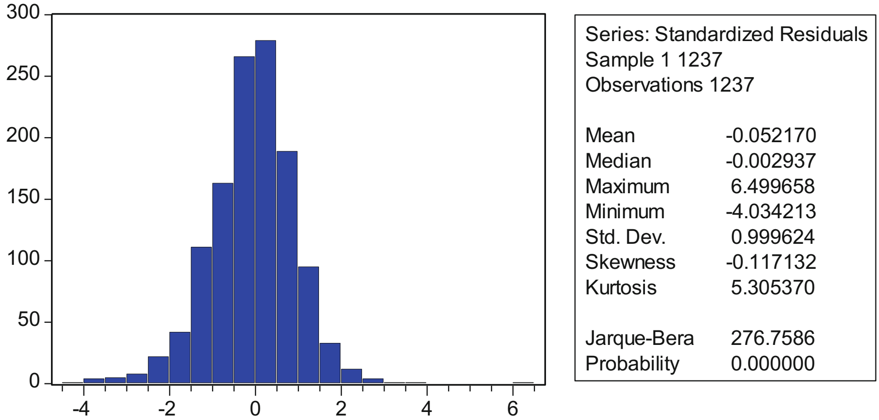


Conditional standard deviation curve of India

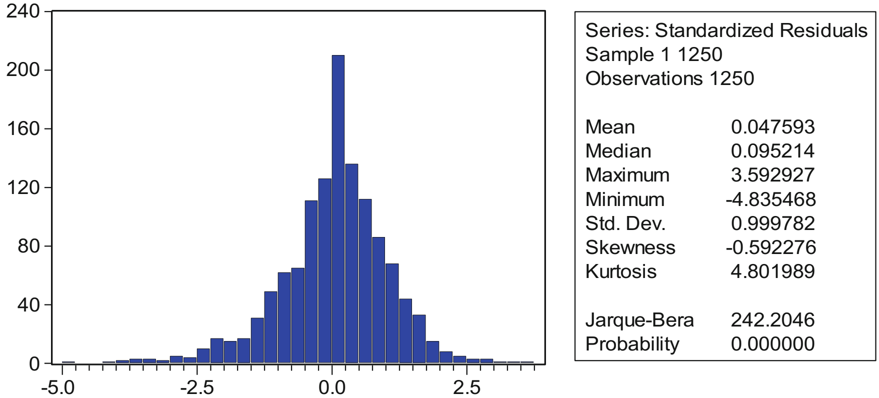


Standardized residual plot of GARCH

SENSEX-INDIA



SCI-CHINA



4 EGARCH Model: Results

4.1 EGARCH Estimates to Justify Volatility Clustering and Leverage Effect

EGARCH model basically models the log of the variance as a function of the lagged log and the lagged absolute error from the regression model. It also allows the response to the lagged error to be asymmetric, so that positive regression residuals can have a different effect on variance than an equivalent negative residual. The results are presented in Table 10 (refer Appendix 2).

We find that all the coefficients are significant. Both in the case of SENSEX and SCI, the asymmetry coefficient γ that is leverage term is negative and statistically different from zero indicating the existence of the leverage effect for the stock market returns during the sample periods. Negative leverage effect implies that the variance goes up more after negative residuals or returns than after positive returns. That means bad news generate more volatility than the good news both in Indian and Chinese stock market during the financial crisis. The only exception we found in case of SCI China is in the first sub-period when the world economy was moving up before the subprime crisis. We found that that although the γ parameter is statistically significant but it is positive. When $\gamma > 0$, it implies that positive impacts are more destabilizing than negative impacts implying good news generate more volatility than bad news. The persistence parameter β is large and continuously increased for SENSEX for three sub-period, implying that variance moves slowly through time or shock became more persistence. But in case of SCI the result was opposite representing a reduction in shock persistence during and after the world recession. The reason may be China has a huge trading surplus with foreign countries, especially the United States. The balance of trade between China and the United States was, by Feb 09, US $34.8 billion in favour of China. The President of the United States in 2009, Barack Obama, had accused China of fixing their exchange rates to take advantage of foreign trade and investment. It may be noted

Table 10 EGARCH estimates to justify volatility clustering and leverage effect in three

	Period-1: 4th March 2006–14th December 2007		Period-2: 17th 2007–1st April 2009		Period-3: 2nd April–31st March 2011	
	SENSEX	SCI	SENSEX	SCI	SENSEX	SCI
ω	−0.373576 (0.0079)	−0.100532 (0.0000)	−0.110166 (0.0653)	0.067374 (0.0000)	−0.212070 (0.0000)	−0.179784 (0.0096)
α	0.322985 (0.0000)	0.134752 (0.0000)	0.152771 (0.0000)	−0.073897 (0.0000)	0.257810 (0.0000)	0.077045 (0.1103)
γ	−0.262083 (0.0000)	0.056456 (0.0000)	−0.125305 (0.0000)	−0.059393 (0.0062)	−0.102913 (0.0007)	−0.120139 (0.0028)
β	0.864514 (0.0000)	0.981920 (0.0000)	0.955006 (0.0000)	0.968566 (0.0000)	0.987444 (0.0000)	0.846203 (0.0000)

that the purchasing power parity in China is equal to about ten times as much as many Western countries. In fact, in 2008 China was ranked the second largest economy in the world using the PPP as a measurement, second only to the US.

4.2 EGARCH Estimate to Justify Spillover from US Markets in Different Sub-periods

In 1993, when Foreign Institutional Investors (FII) were allowed to invest in the Indian equity market, the returns to the domestic investors started becoming increasingly integrated with the ups and downs in the other stock markets of the world. Over time, foreign portfolio investments have increased. Moreover, quite a few Indian listed companies have issued instruments such as American Depository Receipts (ADR) and Global Depository Receipts (GDR) and got their equity shares listed on the US bourses such as NASDAQ and NYSE and European bourses such as LSE. The trend specially gathered momentum during the ICE boom of 1998–2001. The simultaneous listing of a number of large Indian companies in the ICE sector on the Indian bourses and NASDAQ was then expected to generate sympathetic movements or return and volatility 'spillover' across the two markets. These stocks also had significant weight in both the domestic indices.

Hansda and Ray (2002, 2003), Kumar and Mukhopadhyay (2002), Choudhury (2000), Rao and Naik (1990) and Sharma and Kennedy (1977) have studied the spillover between the US and the Indian stock markets. Hansda and Ray (2003) studied the price interdependence of ten Indian companies whose stocks are dually listed, i.e. on the BSE and NSE and the Nasdaq/NYSE. The finding of a bi-directional causality in a vector auto-regression model corroborates the strong correlation between the prices of the dually listed stocks. Furthermore, the impulse responses pattern indicates that a positive shock in the domestic (international) price of a stock gets transmitted in terms of strong positive movement in the international (domestic) price the next trading day. Thus, in addition to stock specific bi-directional causality, the markets are efficient in processing and incorporating the pricing information. Kumar and Mukhopadhyay (2002) employed a two stage GARCH and ARMA-GARCH model to capture the mechanism by which NASDAQ composite daytime returns and volatility have an impact on the conditional mean and the conditional volatility of Nifty overnight returns during the July 1999–June 2001 period. They found that the previous day's daytime returns of both NASDAQ composite and Nifty have significant impact on the Nifty overnight returns on the following day. However, the volatility spillover effects are significant only from NASDAQ composite implying that the conditional volatility of Nifty overnight returns is imported from the US. Choudhury (2000) examined the relationship between SENSEX and Nasdaq returns and found that the correlation is not significant. Similarly, Rao and Naik (1990) examined the inter-relatedness of the US, Japanese, and Indian stock markets using cross-spectral analysis. They concluded that the relationship of the Indian market with international markets is poor reflecting the institutional fact that the Indian economy has been characterized by heavy controls throughout the entire seventies with liberalization measures initiated only in the late eighties. Sharma and Kennedy (1977) examined the price behaviour of the Indian market with US and London markets. They concluded that stocks on the Bombay Stock Exchange obey a random walk and are, in this sense equivalent to the behaviour of stock prices in the markets of advanced industrialized countries like the UK and the US.

In this study, we have applied asymmetrical GARCH models to study the relationship between the US and the domestic markets of India and China represented by SENSEX and SCI. We use the Dow Jones (DJ) indices as the proxy for the US market. We compute the daily return series on Dow Jones in the same way as we have done for SENSEX. This series is then introduced as explanatory variables first in the mean equation and then in the variance equation.

Due to the fact that earlier we have found structural breaks using Chow test and divided the entire period of study in three sub-periods we separately estimated the volatility spillover for these periods. We have introduced Dow Jones with one period lag as explanatory variables into the mean and variance equations estimated separately for Period-1: 4th March 2006–14th December 2007, Period-2: 17th 2007–1st April 2009, Period-3: 2nd April–31st March 2011.

This approach was followed to verify (a) the presence of significant correlation between the two markets and (b) whether the correlation with the US stock markets was any different during the upswing and downswing of the stock market due to recessionary shock. The results of the estimations are presented in Table 11. The results indicate that while Dow Jones exhibits no significant positive correlation with SCI throughout the three sub-periods, in the case of SENSEX only in the first sub-period the correlation was positive. This implies some spillover was taking place before the shock during the first sub-period when SENSEX was moving up to all time high of 20,000. The picture is different in the context of volatility. Positive

Table 11 EGARCH estimates for spillover effect (refer Appendix 3)

	SENSEX EGARCH			SCI EGARCH		
	4/3/2006–14/12/2007	17/12/2007–1/4/2009	2/4/2009–31/3/2011	4/3/2006–14/12/2007	17/12/2007–1/4/2009	2/4/2009–31/3/2011
Mean equation (effect on returns)						
USA DJ	0.160136 (0.0097)	−0.045623 (0.4697)	−0.027055 (0.5087)	0.105099 (0.1536)	0.097534 (0.1371)	−0.091560 (0.0518)
Variance equation (effect on volatility)						
USADJ	0.892383 (0.0000)	0.954788 (0.0000)	0.977461 (0.0000)	0.958284 (0.0000)	0.969752 (0.0000)	0.749673 (0.0000)

correlation is seen between the US indices and SENSEX as well as SCI during all the periods taken into consideration.

The impact of this volatility spillover was quite evident on Indian economy although, initially, it was argued that India would be relatively immune to this crisis, because of the 'strong fundamentals' of the economy and the supposedly well-regulated banking system. But, a crisis of this magnitude was bound to affect globalized economy like India and it did which explains the spillover of volatility from USA. Economy began to slow down from the middle of 2007–2008. Ministry of Commerce report of 03 Aug, 2009 indicated that India registered a fall of 27.7 % in exports for the 9th month in a row in Jun 2009. The imports also dropped by 29.3 %, reflecting a slowdown in domestic consumption. The oil imports in June 2009 plunged by 50.6 % as compared to the same period in 2008. The most immediate effect of that crisis on India was an outflow of foreign institutional investment from the equity market. The financial crisis created a shortage of money supply and India was also facing a credit crunch especially in terms of foreign exchange and the Indian Banking sector and ex-markets were facing tight liquidity situations. This liquidity crisis along with FII sell off forced the Indian Rupee to devaluate from around Rs 40–50/US$.

5 Conclusion

The volatility in the Indian stock market exhibits characteristics similar to those found earlier in many of the major developed and emerging stock markets, viz., autocorrelation and negative asymmetry in daily returns. As for the stationarity of the variance process from the GARCH estimates, it can be observed that required parameter value is equal to 0.99 for SENSEX and 0.98 for SCI. This is less than unity indicating no violation of the stability condition. The sum, however, is rather close to one, which indicates a long persistence of shocks in volatility. The chapter also finds from its EGARCH estimates that, with respect to returns, while Dow Jones exhibits no significant positive correlation with SCI throughout the three sub-periods, in case of the SENSEX at least in the first sub-period the correlation was positive. This implies that some spillover took place before the shock during the first sub-period when SENSEX moved up to all time high of 20,000. The picture is different in the context of volatility. Positive correlation is seen between the US indices and SENSEX as well as SCI during all the period taken into consideration. Although both the countries had suffered from volatility persistence and spillover, the chapter found the impact and incidence was more for India compared to China.

Appendix 1

GARCH estimate for clustering and half-life calculation

1st period India

Dependent variable: RTINDIA				
Sample: 4/03/2006–12/14/2007				
Included observations: 422				
GARCH = $C(2) + C(3) * \text{RESID}(-1)^2 + C(4) * \text{GARCH}(-1)$				
	Coefficient	Std. error	*z*-Statistic	Prob.
C	0.098926	0.029144	3.394374	0.0007
Variance equation				
C	0.026160	0.007093	3.688360	0.0002
$\text{RESID}(-1)^2$	0.180966	0.040497	4.468580	0.0000
GARCH(−1)	0.775294	0.042797	18.11582	0.0000
R-squared	−0.003247	Mean dependent var		0.057862
Adjusted R-squared	−0.010447	S.D. dependent var		0.721531

GARCH 1st period China

Dependent variable: RTCHINA				
Sample: 4/03/2006–12/14/2007				
Included observations: 422				
GARCH = $C(2) + C(3) * \text{RESID}(-1)^2 + C(4) * \text{GARCH}(-1)$				
	Coefficient	Std. error	*z*-Statistic	Prob.
C	0.128874	0.036255	3.554664	0.0004
Variance equation				
C	0.013697	0.004443	3.082556	0.0021
$\text{RESID}(-1)^2$	0.061899	0.013156	4.704891	0.0000
GARCH(−1)	0.924676	0.012104	76.39616	0.0000
R-squared	−0.000406	Mean dependent var		0.145264
Adjusted R-squared	−0.007586	S.D. dependent var		0.814575

GARCH 2nd period India

Dependent variable: RTINDIA				
Sample: 12/17/2007–4/01/2009				
Included observations: 317				
GARCH = $C(2) + C(3) * \text{RESID}(-1)^2 + C(4) * \text{GARCH}(-1)$				
	Coefficient	Std. error	*z*-Statistic	Prob.
C	−0.058058	0.059031	−0.983521	0.3254
Variance equation				
C	0.075185	0.039518	1.902566	0.0571
$\text{RESID}(-1)^2$	0.131614	0.049000	2.686020	0.0072
GARCH(−1)	0.820384	0.060157	13.63744	0.0000
R-squared	−0.000696	Mean dependent var		−0.089802
Adjusted R-squared	−0.010288	S.D. dependent var		1.204988

GARCH 2nd period China

Dependent variable: RTCHINA				
Sample: 12/17/2007–4/01/2009				
Included observations: 317				
GARCH = $C(2) + C(3)$ * RESID(–1)^2 + $C(4)$ * GARCH(–1)				
	Coefficient	Std. error	z-Statistic	Prob.
C	−0.162750	0.028802	−5.650552	0.0000
Variance equation				
C	0.035248	0.003274	10.76595	0.0000
RESID(−1)^2	−0.040575	0.003209	−12.64421	0.0000
GARCH(−1)	1.018870	4.06E-05	25064.94	0.0000
R-squared	−0.001140	Mean dependent var		−0.123300
Adjusted R-squared	−0.010736	S.D. dependent var		1.170177

GARCH 3rd period India

Dependent variable: RTINDIA				
Sample: 4/02/2009–3/31/2011				
Included observations: 498				
GARCH = $C(2) + C(3)$ * RESID(−1)^2 + $C(4)$ * GARCH(−1)				
	Coefficient	Std. error	z-Statistic	Prob.
C	0.050708	0.022293	2.274645	0.0229
Variance equation				
C	0.005644	0.002633	2.143564	0.0321
RESID(−1)^2	0.118561	0.021685	5.467410	0.0000
GARCH(−1)	0.875519	0.022526	38.86757	0.0000
R-squared	−0.000017	Mean dependent var		0.053451
Adjusted R-squared	−0.006090	S.D. dependent var		0.656688

GARCH 3rd period China

Dependent variable: RTCHINA				
Sample: 4/02/2009–3/31/2011				
Included observations: 498				
GARCH = $C(2) + C(3)$ * RESID(−1)^2 + $C(4)$ * GARCH(−1)				
	Coefficient	Std. error	z-Statistic	Prob.
C	0.029746	0.031494	0.944487	0.3449
Variance equation				
C	0.023166	0.011189	2.070427	0.0384
RESID(−1)^2	0.042569	0.014948	2.847911	0.0044
GARCH(−1)	0.905456	0.035288	25.65900	0.0000
R-squared	−0.000043	Mean dependent var		0.025190
Adjusted R-squared	−0.006116	S.D. dependent var		0.699381

Appendix 2

EGARCH estimate For clustering and leverage effect

1st period

Dependent variable: RTINDIA				
Sample: 4/03/2006–12/14/2007				
Included observations: 422				
LOG(GARCH) = *C*(2) + *C*(3) * ABS(RESID(−1)/@SQRT(GARCH(−1))) + *C*(4) * RESID (−1)/@SQRT(GARCH(−1)) + *C*(5) * LOG(GARCH(−1))				
	Coefficient	Std. error	*z*-Statistic	Prob.
C	0.069703	0.026230	2.657399	0.0079
Variance equation				
C(2)	−0.373576	0.063735	−5.861376	0.0000
C(3)	0.322985	0.069004	4.680653	0.0000
C(4)	−0.262083	0.040088	−6.537610	0.0000
C(5)	0.864514	0.022603	38.24711	0.0000
R-squared	−0.000270	Mean dependent var		0.057862
Adjusted R-squared	−0.009865	S.D. dependent var		0.721531

Dependent variable: RTCHINA				
Sample: 4/03/2006–12/14/2007				
Included observations: 422				
LOG(GARCH) = *C*(2) + *C*(3) * ABS(RESID(−1)/@SQRT(GARCH(-1))) + *C*(4) * RESID (−1)/@SQRT(GARCH(−1)) + *C*(5) * LOG(GARCH(−1))				
	Coefficient	Std. error	*z*-Statistic	Prob.
C	0.146602	0.030127	4.866224	0.0000
Variance equation				
C(2)	−0.100532	0.018940	−5.308044	0.0000
C(3)	0.134752	0.027813	4.844927	0.0000
C(4)	0.056456	0.023697	2.382382	0.0172
C(5)	0.981920	0.006318	155.4087	0.0000
R-squared	−0.000003	Mean dependent var		0.145264
Adjusted R-squared	−0.009595	S.D. dependent var		0.814575

2nd period

Dependent variable: RTINDIA				
Sample: 12/17/2007–4/01/2009				
Included observations: 317				
LOG(GARCH) = *C*(2) + *C*(3) * ABS(RESID(−1)/@SQRT(GARCH(−1))) + *C*(4)*RESID(−1)/@SQRT(GARCH(−1)) + *C*(5) * LOG(GARCH(−1))				
	Coefficient	Std. error	*z*-Statistic	Prob.
C	−0.102140	0.059328	−1.721614	0.0851
Variance equation				
C(2)	−0.110166	0.060657	−1.816224	0.0693
C(3)	0.152771	0.075284	2.029272	0.0424
C(4)	−0.125305	0.028927	−4.331755	0.0000
C(5)	0.955006	0.020300	47.04521	0.0000
R-squared	−0.000105	Mean dependent var		−0.089802
Adjusted R-squared	−0.012927	S.D. dependent var		1.204988

Dependent variable: RTCHINA				
Sample: 12/17/2007–4/01/2009				
Included observations: 317				
LOG(GARCH) = *C*(2) + *C*(3) * ABS(RESID(−1)/@SQRT(GARCH(−1))) + *C*(4) * RESID(−1)/@SQRT(GARCH(−1)) + *C*(5) * LOG(GARCH(−1))				
	Coefficient	Std. error	*z*-Statistic	Prob.
C	−0.136118	0.053393	−2.549352	0.0108
Variance equation				
C(2)	0.067374	0.008209	8.207774	0.0000
C(3)	−0.073897	0.005684	−13.00042	0.0000
C(4)	−0.059393	0.021711	−2.735566	0.0062
C(5)	0.968566	0.005465	177.2196	0.0000
R-squared	−0.000120	Mean dependent var		−0.123300
Adjusted R-squared	−0.012942	S.D. dependent var		1.170177

3rd PERIOD

Dependent variable: RTINDIA				
Sample: 4/02/2009–3/31/2011				
Included observations: 498				
LOG(GARCH) = *C*(2) + *C*(3) * ABS(RESID(−1)/@SQRT(GARCH(−1))) + *C*(4) * RESID(−1)/@SQRT(GARCH(−1)) + *C*(5) * LOG(GARCH(−1))				
	Coefficient	Std. error	*z*-Statistic	Prob.
C	0.035054	0.021637	1.620078	0.1052
Variance equation				
C(2)	−0.212070	0.036652	−5.785965	0.0000
C(3)	0.257810	0.041170	6.262093	0.0000
C(4)	−0.102913	0.030323	−3.393880	0.0007
C(5)	0.987444	0.008039	122.8386	0.0000
R-squared	−0.000786	Mean dependent var		0.053451
Adjusted R-squared	−0.008906	S.D. dependent var		0.656688

Dependent variable: RTCHINA				
Sample: 4/02/2009–3/31/2011				
Included observations: 498				
LOG(GARCH) = C(2) + C(3) * ABS(RESID(−1)/@SQRT(GARCH(−1))) + C(4) * RESID(−1)/@SQRT(GARCH(−1)) + C(5) * LOG(GARCH(−1))				
	Coefficient	Std. error	z-Statistic	Prob.
C	0.028783	0.031867	0.903230	0.3664
Variance equation				
C(2)	−0.179784	0.069434	−2.589290	0.0096
C(3)	0.077045	0.048245	1.596959	0.1103
C(4)	−0.120139	0.040140	−2.993006	0.0028
C(5)	0.846203	0.059197	14.29467	0.0000
R-squared	−0.000026	Mean dependent var		0.025190
Adjusted R-squared	−0.008140	S.D. dependent var		0.699381

Appendix 3

EGARCH estimate for Spillover1st sub-period

Dependent variable: RTINDIA				
Sample: 4/03/2006–12/14/2007				
Included observations: 422				
LOG(GARCH) = C(3) + C(4) * ABS(RESID(−1)/@SQRT(GARCH(−1))) + C(5) * RESID(−1)/@SQRT(GARCH(−1)) + C(6) * LOG(GARCH(−1))				
	Coefficient	Std. error	z-Statistic	Prob.
C	0.078013	0.024172	3.227418	0.0012
RTUSA	0.160136	0.061943	2.585198	0.0097
Variance equation				
C(3)	−0.352640	0.083664	−4.214963	0.0000
C(4)	0.314144	0.090026	3.489479	0.0005
C(5)	−0.264987	0.056315	−4.705423	0.0000
C(6)	0.892383	0.029893	29.85266	0.0000
GED parameter	1.341309	0.132189	10.14687	0.0000
R-squared	0.013994	Mean dependent var		0.057862
Adjusted R-squared	−0.000262	S.D. dependent var		0.721531
S.E. of regression	0.721626	Akaike info criterion		1.860797

Dependent variable: RTCHINA				
Sample: 4/03/2006–12/14/2007				
Included observations: 422				
LOG(GARCH) = C(3) + C(4) * ABS(RESID(−1)/@SQRT(GARCH(−1))) + C(5) * RESID(−1)/@SQRT(GARCH(−1)) + C(6) * LOG(GARCH(−1))				
	Coefficient	Std. error	z-Statistic	Prob.
C	0.152740	0.026043	5.864823	0.0000
RTUSA	0.105099	0.073655	1.426903	0.1536
Variance equation				
C(3)	−0.153487	0.051721	−2.967602	0.0030
C(4)	0.192909	0.071931	2.681846	0.0073
C(5)	0.009560	0.049847	0.191793	0.8479
C(6)	0.958284	0.023377	40.99332	0.0000
GED parameter	1.049438	0.095773	10.95762	0.0000
R-squared	0.001252	Mean dependent var		0.145264
Adjusted R-squared	−0.013188	S.D. dependent var		0.814575
S.E. of regression	0.819929	Akaike info criterion		2.221829

2nd sub-period

Dependent variable: RTINDIA				
Sample: 12/17/2007–4/01/2009				
Included observations: 317				
LOG(GARCH) = C(3) + C(4)*ABS(RESID(−1)/@SQRT(GARCH(−1))) + C(5) * RESID(−1)/@SQRT(GARCH(−1)) + C(6) * LOG(GARCH(−1))				
	Coefficient	Std. error	z-Statistic	Prob.
C	−0.105695	0.060773	−1.739173	0.0820
RTUSA	−0.045623	0.063101	−0.723027	0.4697
Variance equation				
C(3)	−0.106505	0.063488	−1.677548	0.0934
C(4)	0.147896	0.079577	1.858535	0.0631
C(5)	−0.129890	0.030174	−4.304656	0.0000
C(6)	0.954788	0.022615	42.21920	0.0000
GED parameter	1.944783	0.245918	7.908260	0.0000
R-squared	−0.004109	Mean dependent var		−0.089802
Adjusted R-squared	−0.023544	S.D. dependent var		1.204988

Dependent variable: RTCHINA				
Sample: 12/17/2007–4/01/2009				
Included observations: 317				
LOG(GARCH) = C(3) + C(4) * ABS(RESID(−1)/@SQRT(GARCH(−1))) + C(5) * RESID(−1)/@SQRT(GARCH(−1)) + C(6) * LOG(GARCH(−1))				
	Coefficient	Std. error	z-Statistic	Prob.
C	−0.096887	0.046488	−2.084100	0.0372
RTUSA	−0.097534	0.065605	−1.486692	0.1371
Variance equation				
C(3)	0.075368	5.38E−08	1400563.	0.0000
C(4)	−0.088226	0.004521	−19.51454	0.0000
C(5)	−0.066091	0.029370	−2.250329	0.0244
C(6)	0.969752	0.005868	165.2748	0.0000
GED parameter	1.462581	0.180404	8.107258	0.0000
R-squared	−0.001916	Mean dependent var		−0.123300
Adjusted R-squared	−0.021307	S.D. dependent var		1.170177

3rd Sub-period

Dependent variable: RTINDIA				
Sample (adjusted): 4/02/2009–3/24/2011				
Included observations: 493 after adjustments				
LOG(GARCH) = C(3) + C(4) * ABS(RESID(−1)/@SQRT(GARCH(−1))) + C(5) * RESID(−1)/@SQRT(GARCH(−1)) + C(6) * LOG(GARCH(−1))				
	Coefficient	Std. error	z-Statistic	Prob.
C	0.040485	0.020647	1.960833	0.0499
RTUSA	−0.027055	0.040936	−0.660907	0.5087
Variance equation				
C(3)	−0.198975	0.055644	−3.575880	0.0003
C(4)	0.222863	0.059845	3.723999	0.0002
C(5)	−0.115798	0.041855	−2.766615	0.0057
C(6)	0.977461	0.015055	64.92544	0.0000
GED parameter	1.362674	0.090024	15.13686	0.0000
R-squared	−0.003205	Mean dependent var		0.053993
Adjusted R-squared	−0.015590	S.D. dependent var		0.659994

Dependent variable: RTCHINA				
Sample: 4/02/2009–3/31/2011				
Included observations: 498				
LOG(GARCH) = C(3) + C(4) * ABS(RESID(−1)/@SQRT(GARCH(−1))) + C(5) * RESID(−1)/@SQRT(GARCH(−1)) + C(6) * LOG(GARCH(−1))				
	Coefficient	Std. error	z-Statistic	Prob.
C	0.070085	0.026711	2.623857	0.0087
RTUSA	−0.091560	0.047075	−1.944979	0.0518
Variance equation				
C(3)	−0.219788	0.110867	−1.982449	0.0474
C(4)	0.016437	0.086411	0.190224	0.8491
C(5)	−0.202779	0.077343	−2.621820	0.0087
C(6)	0.749673	0.109840	6.825106	0.0000
GED parameter	1.262590	0.114267	11.04950	0.0000
R-squared	−0.003006	Mean dependent var		0.025190
Adjusted R-squared	−0.015262	S.D. dependent var		0.699381

References

Aggarwal R, Tandon K (1994) Anomalies or illusions? Evidence from stock markets in eighteen countries. J Int Money Finan 13(1):83–106

Bollerslev T (1986) Generalised autoregressive conditional heteroskedasticity. J Econom 31 (3):307–327

Booth GG, Martikainen T, Tse Y (1997) Price and volatility spillovers in Scandinavian stock markets. J Bank Finan 21(6):811–823

Box GEP, Jenkins GM (1976) Time series analysis: forecasting and control, revised edn. Holden-Day, California

Box G, Pierce D (1970) Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. J Am Stat Assoc 65:1509–1526

Chaudhury SK (1991) Seasonality in share returns: preliminary evidence on day of the week effect. Chartered Accountant (India) 40(5):107–109

Fama EF (1965) The Behaviour of Stock Market Prices. J Bus 38(1):34–105

Garman M, Klass M (1980) On the estimation of security price volatilities from historical data. J Bus 53:67–78

Hansda SK, Ray P (2002) BSE and Nasdaq: Globalisation, Information Technology and Stock Prices. Economic and Political Weekly, February 2, 459–468

Hansda SK, Ray P (2003) Stock Market Integration and Dually Listed Stocks: Indian ADR and Domestic Stock Prices. Economic and Political Weekly, February 22, 741–753

Kumar KK, Mukhopadhyay C (2002) A Case of US and India. Paper published as part of the NSE Research Initiative. http//:www.nseindia.com

Lamoureux Christopher G, Lastrapes William D (1990) Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects. J Financ 45(1):221–229

Ljung G, Box G (1978) On a measure of lack of fit in time series models. Biometrika 66:67–72

Lo AW, MacKinlay AC (1988) Stock market prices do not follow random walks: evidence from a simple specification test. Rev Financ Stud 1:41–66

Mandelbrot B (1963) The Variation of Certain Speculative Prices. J Bus 36(4):394–419
Miller MH (1991) Financial innovations and market volatility. Blackwell, pp 1–288
Parkinson M (1980) The extreme value method for estimating the variance of the rate of return. J Bus 53:61–65
Poterba James M, Summers LH (1986) The Persistence of Volatility and Stock Market Fluctuations. Am Econ Rev 1142–1151
Rao BSR, Naik U (1990) Inter-Relatedness of Stock Markets: Spectral investigation of USA, Japanese and Indian Markets – A Note. Arth Vignana 32(3&4):309–321
Sharma JL, Kennedy RE (1977) A Comparative Analysis of Stock Price behavior on Bombay, London and New York Stock Exchange. J Financ Quant Anal 31(3):391–413

Imbalances, Local and Global, and Policy Challenges in the Post-Crisis World

Soumyen Sikdar

1 Introduction

The global financial meltdown of 2008–2009 following the housing market collapse in the USA is surely the most severe crisis in capitalism since the Great Depression of the 1930s. Christened the Great Recession, it has naturally caused a great deal of rethinking by economists and policy makers on a great variety of fundamental issues ranging from globalization, deregulation and financial management by central banks to the probity of highly paid company CEOs and the reliability of international credit rating agencies. Interesting, but not surprising, is the resurfacing of the ideas of Marx and Keynes. Symptomatic of the trend is the immense stir created by Thomas Piketty's book, Capital in the Twenty-First Century. Complacence and certitude have received rude jolts across the world and the need has emerged to think afresh on numerous policy fronts. India's integration with the rest of the world is still comparatively low. Yet, she could not escape unhurt. Prompt policy measures were successful in cushioning the shock, the worst could be averted but the economy is yet to recover fully and get back to the track of pre-crisis performance in terms of growth.

Though it has no obvious connection with the global catastrophe; at the national level too India has experienced a major political change. For the first time in many years, a party has been voted to power with overwhelming popular support. The government is no longer shackled by the stringent compulsions of coalitional politics. In the past coalition management (the perennialcurse of Indian democracy) and legislative logjam often led to policy paralysis with all its seriously adverse consequences. This constraint no longer obtains in India. So, the government has more freedom and flexibility to address the issues that have attained prominence in the post-crisis global scenario.

S. Sikdar (✉)
Indian Institute of Management Calcutta, Kolkata, India
e-mail: soumyen1956@gmail.com

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_21

In National People's Congress in March 2007, the Chinese Prime Minister Wen Jiabao expressed concern that China is increasingly becoming 'unbalanced, uncoordinated and unstable.' These alarming adjectives capture India's situation equally well. They merit close attention from those in charge of making policies.

Now we take a look at a number of imbalances that have emerged (and are growing) both at global and local levels. Long run stability of the world's economic order is seriously under threat as a consequence of these developments. The stiff challenge that the crisis has thrown in the face of central bankers, a very important and influential policy making group, is also taken up. It is argued that, instead of targeting inflation exclusively they should adopt a more balanced position while deciding policy. In other words, central bank objective functions should have a more balanced distribution of weights. Clearly, this particular 'imbalance' is not of the same kind as the other ones taken up in the discussion such as sectoral or distributional imbalances, but we consider it important enough (as source of critical policy failure) for inclusion in the context of the current global crisis.

The chapter is organized as follows. The next section looks at one major manifestation of global imbalance, namely, overdependence on the performance of a single country, the USA. This is followed by a look at macro imbalances on the home front. Attention is drawn to rising inequality in income and wealth, both in India and the world at large in Sect. 4. This is important because imbalance and iniquity in the sharing of the fruits of 'progress' calls into question the legitimacy of the entire programme of growth under free market capitalism. Section 5 discusses one vital failure in policy formulation, namely, the extremely lopsided view that the task of central banking begins and ends with inflation control. There is urgent need to restore balance in this sphere too. The final section concludes.

2 Global Imbalance

The global economy is too much dependent on a particular country, namely, the USA. It is American consumption that has become the motor of the world's prosperity. To give an idea of the motor's power, real consumption over the period 1993–2015 grew in the USA at an average rate of 4 %, which is no less than three times the consumption growth of Europe and Japan combined. American consumption-income ratio reached the value of 72 % in 2007, an all time record. In absolute terms, for the sake of comparison, the values of consumption expenditure in that year were $9.5 trillion for the USA, $1 trillion for China and $650 billion for India.

The notion of Global Decoupling was greatly in vogue for a number of years preceding the recent catastrophe. The putative decoupling was between the USA and a collection of Asian and Latin American countries including India, Thailand, Malaysia and Vietnam from Asia (but not China) and Brazil from South America. For these countries, the importance of the USA as a trade partner has been gradually decreasing and it was hoped that they would be able to function as a relatively autonomous group in the near future. But the hope turned out to be illusory as consumption in the USA crashed after the crisis and one country after another was pulled down into the abyss. On more than one occasion, Chairman Ben Bernanke of the Federal Reserves had earlier pointed to the 'saving glut' of Asia as a prominent sign of structural imbalance in the world. Actually, it is the consumption frenzy of the Americans which should be blamed in this context, rather than the so called Asian saving glut.

The message is clear. The degree of dependence of the rest of the world on any single country must not be allowed to rise beyond a limit.

3 Local Imbalances

The Indian economy could escape the Great Recession with relatively minor damage precisely because its linkage with the American economy is not very strong. After the worst year of 2008–2009 recovery has been fairly quick. However, its own performance over the past quarter century reveals an extremely undesirable trait. Growth has been very unbalanced between the three major sectors- agriculture (primary), industry (secondary) and services (tertiary). Agriculture has suffered from secular stagnation, manufacturing barely manages to limp along, while services have grown at a very high rate. In the course of the last two decades, agriculture's share has been declining continuously. It h now stands at 16 %, while manufacturing has managed to maintain its share at an average of 20–24 % over 1980–1981 to 2009–2010. This is dwarfed by Thailand (36 %), South Korea (32 %), China (45 %) and Taiwan (30 %). India's share of global manufacturing is a tiny 2.2 %, compared with China's 18.9 %. Another glaring imbalance in the Indian economy is the fact that although agriculture contributes only 22 % of GDP, it engages close to 50 % of the total workforce. Over the period 1980–2000, agriculture's share (in percentage) in total employment declined from 68 to 59 in India. Over the same period, China managed to bring it down from 70 to 48.

The services component of India's national income has swelled and swelled and at present commands a share of nearly 60 %. Actually, our strong economic performance in recent (pre-crisis) years has been powered mostly by the growth in the export of services.

India's share in total global service exports is now almost 4 %. There was a marked slowdown in 2009–2010 following the severe global crisis, but the damage was smaller than in merchandise exports. Recovery also was very quick and service exports are back at their pre-crisis levels. Given that it will be very difficult to compete with China in the field of manufacturing exports, our growing success as service exporter indeed augurs well for the future of the economy.

Two questions have arisen in this context. First, is India following an 'abnormal' trajectory deviating from the Kuznetsian sequence of primary-secondary-tertiary growth? The simple answer is that there is no fixed and predetermined development path that every country is destined to follow. The Indian experience of service-led growth may well become a unique model for the emerging market economies.

The second question is much more important. Is the current pattern of sectorally unbalanced growth sustainable over time? If it is not, then the economic impetus will soon peter out and we may slip back into the old days of stagnation or very sluggish growth.

Is it likely that external demand for our services may dry up in the near future? This is not a serious threat because the trend of outsourcing by corporates in the high wage advanced countries is not likely to be reversed soon. Services account for more than 60 % of global GDP at present and have high income elasticity. With global income recovering after the severe slump, demand for services, both as production input and as final consumption, is likely to resume its steady growth.

Slackening of external demand may, however, arrive in the form of demand switch away from India. This will happen if there is a drop in our global competitiveness. Alarming signs are already noticeable. One factor contributing to decline in competitiveness may be relative wage inflation. That is, wages in our service sector rising faster than those of our competitors. And a major factor behind the rise in the remuneration of skilled workers in India (in real terms) may be the widening gap between demand and supply in the labour market. There are many studies that draw attention to the various problems of skill development in India, highlighting in the process the mismatch between industry needs and the output of educational institutions (Murugaia et al. 2014). The literature also reveals that China, Vietnam, Indonesia and the Philippines are racing ahead of us in some crucial areas of skill formation. (For a comparison with China, see Asuyama 2009.) Our market share in business services may begin to shrink if we fail to maintain a steady rise in the supply of adequately trained workers. In that event growth of our GDP will take a nasty hit.

There is, therefore, an urgent need to bring our exports out of the very narrow spectrum within which they are confined at present. There is considerable scope for diversifying exports away from services in general and business services in particular into other types of activities. For that to happen, our manufacturing cum export oriented infrastructure has to be expanded and improved. The incentive regime still favours the domestic market by and large and protection of inefficient industries still persists, though at a lower level compared to the pre-reform days (Alfaro and Chari 2013). Concerted attempt must be made to design and implement better policies and reverse the serious disparity in growth between the three major segments of the economy.

Imbalance is also growing among the states of India. Both FDI and BPO activities by MNCs show a marked regional bias, with some states (Gujarat, Karnataka, Maharashtra and Tamil Nadu) gaining disproportionately more than the others. In the context of regional divergence, Maiti and Marjit (2010) examined whether over the period 1980–2004 greater openness has had an equalizing impact or not. Their major conclusion is that more open states grew faster by 1–1.5 % per annum. States that were able to change their production structure towards export production showed greater increment in growth. And this ability was strongly correlated with the quality of institutions and the prevailing investment climate.

4 Inequality in Income and Wealth

Another very serious manifestation of imbalance is that in the distribution of income and wealth, both within and across countries. Piketty's fundamental research (Piketty 2014) into the evolution of inequality in capitalist countries and the contributory factors has induced a great upsurge of interest in this problem. Within the developing world, the gap between the rich and poor is growing steadily both in India and China, though possibly at a faster rate in the latter. Our recent export success has been mostly confined to the IT sector which is intensive in the use of skilled labour and which does not have strong linkages with the rest of the economy. Its concentration in particular areas of the country has contributed to a widening of regional disparity. If the benefits of globalization continue to be enjoyed so unequally, the whole growth process runs the risk of being disrupted by social disharmony and upheaval.

Improvement in labour's share in national income in India has been held back by the stagnant (or declining) share of labour intensive products in manufacturing and exports. Das et al. (2009) identified thirty-one four digit industries, such as food and beverages, readymade garments and apparels, textiles products and furniture manufacturing as labour intensive. Gross value added (GVA) in these industries between 1990–1991 and 2003–2004 averaged only 13.8 % of the total organized sector manufacturing GVA. There are reasons to believe that since then the overall production structure of the economy has shifted further towards higher capital intensity. Some of the fastest growing sectors over the last couple of decades have been two, three and four wheeler vehicles, auto parts, software, telecommunications equipment, petroleum refining, pharmaceuticals and finance. None of these offers much scope for employment to low-skilled workers.

Our export composition also has increasingly tilted towards capital and skilled-labour intensive components. In manufacturing engineering goods and petroleum products continue to show the largest expansion, while the share of readymade garments, the most low-skilled labour absorbing product, fell from 12 to just 6 % over 1990–1991 to 2007–2008 (Panagariya 2011). Unless these tendencies are reversed or moderated, distribution of factor earnings will continue to be characterized by rising inequality.

5 Wake up Call for Central Bankers

The latest crisis originated in the financial sector of the USA. It is now clear that flawed supervision and failure of regulation by the Federal Reserve Board (Fed) played a major role behind the catastrophe. That the Indian economy could escape relatively unhurt is in no small measure due to the prudential stance of the RBI consistently maintained over the years.

How could the Fed (and other central banks of the advanced countries to a lesser extent) be remiss on such a big scale for so long? Apart from the possibility of outright insensitivity (or nefarious nexus with Wall Street, which cannot be ruled out), what is at fault is a fundamentally flawed view of the duty of the monetary authorities. It is the view that inflation control is the one and only relevant policy objective of the central bank. This is an exemplar of serious lopsidedness or lack of balance in policy formulation and setting of priorities.

Inflation Targeting acquired prestige in the USA after the spectacular (and much publicized) victory of Fed Chairman Paul Volcker over the double digit inflation raging in the 1970s in the country. Subsequently, it became enshrined as the official policy of the Fed. Since 1988, it has been explicitly adopted by several countries. New Zealand was the first to do so in 1989, followed by Canada and Israel (1991), the UK (1992), Sweden, Finland and Australia (1993) and Spain (1994). While the earlier policy of monetary targeting had been drawn up after intensive academic debate and discussion, inflation targeting was adopted ad hoc from the American model. 'Stability-oriented monetary policy' became virtually identified with keeping inflation under tight control. In the words of Ben Bernanke, another Fed Chairman, 'Low, stable inflation is monetary policy's primary long run goal.' As it happened, it quickly became the primary short run goal also. In the singleminded concentration on price stability the vital supervisory role of central bankers was completely ignored. In particular, the evil of unemployment ceased to be a matter of concern altogether. To quote Bernanke and his co-authors of *Inflation Targeting*, 'Contrary to what was believed 30 years ago, it appears that the benefits of expansionary policies (such as lower unemployment) are largely transitory, whereas the costs of expansionary policies (primarily the inefficiencies associated with higher inflation) tend to be permanent'. (Bernanke et al. 2001) This is the legacy of a strong belief in the idea of the natural rate of output (and employment) towards which an unregulated economy tends automatically to gravitate. (For a cogent critique of 'the natural rate hypothesis' of New Classical Economics see Akerlof 2002.)

It is now widely recognized that the market-friendly low interest policy stance of the Fed maintained quarter after quarter under the stewardship of Alan Greenspan was responsible in a big way for fueling and sustaining the asset bubble in the USA. Following the warnings of Robert Shiller who has intensively studied the behaviour of stock markets over long periods, Greenspan talked of 'irrational exuberance' in asset markets in a speech of December 1996. The anticipation that the Fed would initiate a tight money policy led stock markets around the world to fall sharply. In reaction, Greenspan immediately backed off believing that increase in interest rates will do significant damage to the economy. But he had other tools at his disposal. He could have increased capital requirements for financial institutions, limiting their capacity to borrow, take steps to check widespread fraud in the issuance of mortgages and insisted on greater transparency in the collateralized debt obligations (CDOs) and other novel products of innovative financial engineering. Believing in the power of the Invisible Hand of the free market to achieve efficiency unaided, he took none of these measures. However, after the catastrophic bursting of the bubble he admitted that there was indeed a serious policy mistake. But it was too late to undo the damage already done.

Under the Fed's sustained easy money policy there was hardly any attempt to control the unnatural and clearly unsustainable developments in the financial market and restore orderly conditions there. The crisis has once again highlighted the wisdom that without going for heavy-handed intervention, central banks should constantly monitor developments in the asset markets and use the information as an essential input in monetary policy formulation. Everything's fine because inflation is close to 2 %, this patently absurd view should be given up once and for all. This is possibly the most important lesson for monetary authorities to come out of the colossal policy failure. Actually, in 2013 some members of the Fed including Janet Yellen, who would succeed Bernanke in 2014, began to call for more attention to the poor employment situation. One can only hope that this concern will continue to share equal weight with inflation control in future policy formulations of the Fed.

Over the recent past, RBI's policy too has been showing a tendency to drift towards inflation targeting. Encouraging signs of change in favour of a more balanced approach have begun to appear.

6 Conclusion

The most severe economic crisis since the Great Depression has spawned an immense number of questions about the viable functioning of the global economy driven by the forces of unregulated capitalism. The path to the collapse originated in an ideological turn in the 1970s that blindly sought to eliminate the government rather than reform it. The influence of this biased ideology, on the thinking of policy makers, the world over is profoundly responsible for what has happened in the USA and the rest of the world. The rude shock seems to have induced a change in the mindset of policymakers including central bankers. It is now clear that there can be no substitute for intelligent governance freed of ideology. Financial markets, in particular, must be carefully supervised and regulated. Laissez faire in this sector is a recipe for disaster.

Over the past quarter century or more many glaring imbalances, sectoral, regional, distributional have been steadily on the rise across the globe. Unless effective measures are taken to check or reverse them quickly, growth will inevitably become unsustainable and with high probability the entire superstructure will come crashing down once more.

References

Akerlof G (2002) Behavioral macroeconomics and macroeconomic behavior. Am Econ Rev 92:411–433

Alfaro L, Chari A (2013) Reforms and the competitive environment. In: Bhagwati J, Panagariya A (eds) Reforms and economic transformation in India. Oxford University Press, New Delhi

Asuyama Y (2009) Skill formation systems of China and India. In: Moriki O, Kimura K (eds) Comparative study on industrial development process in China and India. Interim Report, Institute of Developing Economies, Tokyo

Bernanke B, Laubach T, Mishkin F, Posen A (2001) Inflation targeting: lessons from the international experience. Princeton University Press, New Jersey

Das DK, Wadhwa D, Kalita G (2009) The employment potential of labour intensive industries in India's organized manufacturing. ICRIER Working Paper 236, June

Maiti D, Marjit S (2010) Regional openness, income growth and disparity among major Indian states during 1980–2004. Institute of Economic Growth Working Paper 304

Murugaia S, Kubendran V, Vimala A (2014) Skills mismatch between industry needs and institutions output-challenges for higher education in India. Indian J Appl Res 4, 9, September

Panagariya A (2011) Avoiding lopsided spatial transformation. In: Ghani E (ed) Reshaping tomorrow. Oxford University Press, New Delhi

Piketty T (2014) Capital in the twenty-first century. Harvard University Press, Boston

Testing Non-linearity in Emerging and Developed Markets

Kousik Guhathakurta, Basabi Bhattacharya and A. Roy Chowdhury

Abstract Behaviour of non-linearity in stock markets of developed countries has been empirically evidenced for quite some time. Some few studies examine non-linearity of emerging stock markets as well. A comprehensive overview of stock market behaviour in terms of non-linearity is undertaken in this study where several non-linearity test approaches have been applied on both the developed and emerging country groups. For the purpose we have selected seven stock exchanges from developed markets (UK, Germany, Australia, New Zealand, Hong Kong, France and Japan) and seven from emerging markets (Hungary, Chile, Mexico, Malaysia, Taiwan, China and India). Taking their daily close data covering a period from January 2005 to July 2011, we have investigated the presence of nonlinearity in the data sets and tried to find out whether there is any difference amongst them in this regard. We have first used the test method developed by Brock, Dechert and Scheinkman (BDS) and examine the non-linearity feature in each of the time series. To reinforce our findings, we have conducted the White's Neural Network tests on the same data set. Additionally, we have also performed the Keenan's test for non-linearity. Another popular non-linear test is the Hinich bi-spectrum test, which involves estimating the bi-spectrum of the observed time series. We have further used this test to find out whether it detects non-linearity in these time series. Another linearity test for time series was introduced based on concepts from the

K. Guhathakurta (✉)
Finance & Accounting, Indian Institute of Management Indore,
Prabandh Shikhar, Rau-Pithampur Road, Indore, Madhya Pradesh, India
e-mail: kousikg@iimidr.ac.in

B. Bhattacharya
Department of Economics, Jadavpur University, Kolkata 700032,
West Bengal, India
e-mail: basabi54@gmail.com

A. Roy Chowdhury
High Energy Physics Division, Department of Physics, Jadavpur University,
Kolkata 700032, West Bengal, India
e-mail: arc.roy@gmail.com

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_22

theory of neural networks. Teräsvirta et al. developed its power fully. We have applied this Teräsvirta Neural Network test as a final reinforcement of our findings.

Keywords Stock market · Non-linearity · BDS test · White's neural network test · Keenan's test · Hinich bi-spectrum test · Teräsvirta neural network test

1 Introduction

It is an accepted fact that financial economics has been dominated over the past decades by linear paradigm, which assumes that economic time series conform to linear models or can be well approximated by a linear model. For example, empirical tests of market efficiency, purchasing power parity, tests of stationarity, cointegration, causality and many of the empirical models of asset pricing have implicitly assumed that the underlying dynamics are in linear form or can be made linear by a simple transformation.

However, since economic and financial systems are observed to go through both structural and behavioural change, simple linear time series models run the risk of leaving certain aspects of economic and financial data unexplained. Therefore, it is important to understand that different time series models may be required to explain the empirical data (Pesaran and Potter 1993; Campbell et al. 1997; Barnett and Serletis 2000). There exist studies showing that financial time series exhibit non-linear dependencies (Hsieh 1989, 1991; Scheinkman and LeBaron 1989; De Grauwe et al. 1993; Abhyankar et al. 1995; Steurer 1995; Brooks 1996; Barkoulas and Travlos 1998; Opong et al. 1999). Recognition of non-linearity in time series has given a new direction to empirical research. In the words of Campbell et al. (1997, p. 467), "A natural frontier for financial econometrics is the modelling of non-linear phenomenon". Testing for non-linearity has received extreme importance in the financial econometrics literature in recent years, though the focus is on financial markets of developed countries. Hinich and Patterson (1985) were the first among others who established non-linear dependence in NYSE stock returns. In principle, testing for non-linearity can be viewed as general test of model adequacy for linear models (Hinich and Patterson 1989) and it has been argued that if the underlying generating process for a time series is non-linear in nature, then it would be inappropriate to employ linear methods. For instance, most of the widely applied statistical tests like the unit root or stationary tests, the Granger causality test and the cointegration test are all built on the basis of linear autoregressive model. Taylor and Peel (1997) and Sarno (2000), amongst others, illustrated that the adoption of linear stationarity tests are inappropriate in detecting mean reversion if the true data generating process is in fact a stationary non-linear process. On the other hand, the Monte Carlo simulation evidence in Bierens (1997) indicated that the standard linear cointegration framework presents a mis-specification problem when the true nature of the adjustment process is non-linear and the speed of adjustment varies with the magnitude of the disequilibrium. For the purpose, one would have to resort to empirical methods like non-parametric cointegration test due to Bierens (1997), non-linear stationarity tests (Sarno 2001; Chortareas et al. 2002; Kapetanios et al. 2003) and non-linear causality test (Baek and Brock 1992) etc. Therefore, in empirical research testing for non-linearity assumes importance as a preliminary diagnostic tool to determine the nature of the data generating process.

Over the past few decades, numerous studies have documented the existence of non-linear dependencies in exchange rates returns series (Hsieh 1989; De Grauwe et al. 1993; Steurer 1995; Brooks 1996). The stock markets have also attracted the attention of researchers with substantial evidence supporting the presence of non-linearity in stock returns series (Scheinkman and LeBaron 1989; Hsieh 1991; Abhyankar et al. 1995, 1997; Barkoulas and Travlos 1998; Opong et al. 1999; Willey 1992; Lee et al. 1993; Pagan 1996; Blasco et al. 1997; Lima 1998; Dahl and Nielson 2001).

However, much of this evidence has been drawn from the widely traded financial markets of well-developed countries. For emerging markets, Sewell et al. (1993) provided evidence of non-linearity. Similar other studies which may be cited are Cinko (2002) for Turkey, Scheicher (1996) for Vienna, Afonso and Teixeira (1998) for Portugal, Seddighi and Nian (2004) for China, Sharma and Panagiotidis (2005) for Greece and Dorina and Simina (2008) for eight emerging economies (Romania, Hungary, Czech Republic, Lithuania, Poland, Slovakia, Slovenia, Turkey), and Hassan et al. (2003) for Kuwait and among the recent ones Lim and Brooks (2009) for China. In the Indian context, study by Hiremath and Kamaiah (2010) and Bhattacharya and Sensarma (2013) established non-linear dependence in stock returns.

This study investigates the behaviour of the stock market indices in terms of non-linearity for fourteen countries selected both from the developed and the emerging group. Since non-linearity may take place in many ways, for each of these countries several non-linearity tests have been applied to ensure consistency of results across different test approaches and different country groups. In this respect, the paper provides a broad-based comprehensive overview of the non-linear behaviour of the stock markets not considered earlier in the literature.

2 Data Set

For our study, we selected seven stock exchanges from developed markets (UK, Germany, Australia, New Zealand, Hong Kong, France and Japan) and seven from emerging markets (Hungary, Chile, Mexico, Malaysia, Taiwan, China and India) as listed below. We considered their daily closing value in the period January 2005 to July 2011. We also performed the tests on their daily log return time series (Table 1).

To reinforce our findings we also conduct the White's neural Network (Lee et al. 1993) tests on the same data set. Lee et al. (1993) developed new test, the neural network test for neglected non-linearity. The neural network test is based on the approximating ability of neural network modelling techniques recently developed by cognitive scientists. This test is a Lagrange Multiplier test that statistically determines whether adding 'hidden units' to the linear network would be

Table 1 List of stock indices

Country	Stock exchange	Index
Developed		
United Kingdom	London stock exchange	FTSE
Germany	Frankfurt stock exchange	DAX
Australia	Australian securities exchange	AOI
New Zealand	New Zealand stock exchange	NZX
Hong Kong	Hong Kong stock exchange	HangSeng
France	Paris stock exchange	FCHI
Japan	Tokyo stock exchange	NIKKEI
Emerging		
Hungary	Budapest stock exchange	BUX
Chile	Santiago stock exchange	IPSA
Mexico	Mexico stock exchange	MXX
Malaysia	Kuala Lumpur stock exchange	KLSE
Taiwan	Taiwan stock exchange	TWII
China	Shanghai and Shenzhen stock exchanges	CSI
India	National stock exchange	NIFTY

advantageous. This is a powerful test and can be used to determine non-linearity in a data set quite clearly. This test is further developed and investigated by Teräsvirta et al. (1993). They are compared by simulation with those of a Lagrange multiplier (LM) type test that they derive from the same single-hidden-layer neural network model. The auxiliary regression of their LM type test is a simple cubic 'dual' of the Volterra expansion of the original series, and the power of the test appears superior overall to that of the other test.

We performed all the tests using MATLAB and R software.

3 Non-linearity Tests and Results

3.1 Non-linearity

In the literature, there is no generally agreed definition for 'non-linearity'. From the definition given by De Grauwe et al. (1993, p. 244), a system $X_t = h(\Omega_t, \alpha)$, is called a non-linear system if it is not possible to regenerate X_t by one linear model:

$$X_t = \sum_{i=0}^{\infty} \gamma_i\, \varepsilon_{t-i} \text{ and } \varepsilon \text{ is white noise and}$$

$$\sum_{i=0}^{\infty} \gamma_i \text{ is such that} \sum_{i=0}^{\infty} |\gamma_i| < \infty$$

According to De Grauwe et al. (1993), the definition of non-linearity stems from the negation of linearity. This leaves a lot of other possibilities open for a so-called non-linear system. For example, Hsieh (1989) divided the realm of non-linear dependencies into three categories. Additive non-linearity, also known as non-linear-in-mean, enters a process through its mean or expected value, so that each element in the sequence can be expressed as the sum of zero-mean random element and a non-linear function of past elements1. With multiplicative non-linearity, or non-linear-in-variance, each element can be expressed as the product of a zero-mean random element and a non-linear function of past elements, so that the non-linearity affects the process through its variance. The final category is known as hybrid dependence, in which non-linearity enters through both the mean and the variance.

Tsay (2005) has considered non-linearity in the following way.

Let us consider a univariate time series x_t, which, for simplicity, is observed at equally spaced time intervals. We denote the observations by $\{x_t | t = 1 \ldots T\}$, where T is the sample size. A purely stochastic time series x_t is said to be linear if it can be written as

$$x_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i} \tag{1}$$

where μ is a constant, ψ_i are real numbers with $\psi_0 = 1$, and $\{a_t\}$ is a sequence of independent and identically distributed (iid) random variables with a well defined distribution function. We assume that the distribution of at is continuous and $E(a_t) = 0$. In many cases, we further assume that $\mathrm{Var}(a_t) = \sigma_a^2$ or, even stronger, that at is Gaussian. If $\sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2 < \infty$, then x_t is weakly stationary (i.e., the first two moments of x_t are time-invariant). Any stochastic process that does not satisfy the condition of Eq. () is said to be non-linear. The prior definition of non-linearity is for purely stochastic time series. One may extend the definition by allowing the mean of x_t to be a linear function of some exogenous variables, including the time index and some periodic functions.

Under the null hypothesis of linearity, residuals of a properly specified linear model should be independent. Any violation of independence in the residuals indicates inadequacy of the entertained model, including the linearity assumption. This is the basic idea behind various non-linearity tests. In particular, some of the non-linearity tests are designed to check for possible violation in quadratic forms of the underlying time series.

There is no single test that dominates the others in detecting non-linearity because non-linearity may occur in many ways (Tsay 2005). Keeping this in mind, we use different tests to examine whether there is any presence of non-linearity in the data.

3.2 BDS Test

We first use the test method developed by Brock, Dechert and Scheinkman (BDS) (Brock et al. 1996) and test for non-linearity in each of the time series. The test does not provide a direct test for non-linearity because the sampling distribution of the BDS test statistic is not known, either in finite samples or asymptotically, under the null hypothesis of non-linearity. In the BDS test, initially the linear structure in the time series is removed by de-trending or first differencing. The remaining residuals are used to estimate the BDS statistic and the remaining dependence and the presence of omitted non-linear structure is detected. The BDS tests the null hypothesis that the remaining residuals are independent and identically distributed (iid.). Rejection of null hypothesis implicates the presence of non-linearity.

The BDS test well captures the approach of critical transitions in the residual time series triggered by strong non-linear responses. The data set considered here covers the most recent financial crisis episode spread across different stock markets. Though the BDS test can, in principle, be applied within rolling windows to detect increasing non-linearity in a time series approaching transition, this has not been used in the paper. The time period covers the entire turmoil period encompassing the build up pressure heading the crisis, the crisis period and the aftermath of the crisis.

The basic idea of the BDS test is to make use of a "correlation integral" popular in chaotic time series analysis. Given a k-dimensional time series X_t and observations $\{X_t\}_{t=1}^{T_k}$, define the correlation integral as

$$C_k(\delta) = \lim_{T_k \to \infty} \frac{2}{T_k(T_k - 1)} \sum_{i<j} I_\delta(X_i - X_j) \tag{2}$$

where $I_\delta(u,\ v)$ is an indicator variable that equals one if $\|u - v\| < \delta$, and zero otherwise, where $\|.\|$ is the supnorm. The correlation integral measures the fraction of data pairs of $\{X_t\}$ that are within a distance of δ from each other. Next a time series x_t is considered. k-dimensional vectors $X_t^k = (x_t, x_t + 1, \ldots, x_t + k - 1)$ which are called k-histories are constructed. The idea of the BDS test is as follows. A k-history is treated as a point in the k-dimensional space. If $\{x_t\}_{t=1}^{T}$, are indeed iid random variables, then the k-histories $\{X_t\}_{t=1}^{T_k}$, should show no pattern in the k-dimensional space. Consequently, the correlation integrals should satisfy the relation $C_k(\delta) = [C_1(\delta)]^k$. Any departure from the prior relation suggests that x_t are not iid. As a simple, but informative example, let us consider a sequence of iid random variables from the uniform distribution over [0, 1]. Let $[a, b]$ be a subinterval of

$[0, 1]$ and consider the "2-history" $(x_t, x_t + 1)$, which represents a point in the two-dimensional space. Under the iid assumption, the expected number of 2 histories in the subspace $[a, b] \times [a, b]$ should equal the square of the expected number of x_t in $[a, b]$. This idea can be formally examined using sample counterparts of correlation integrals. We define

$$C_l(\delta, T) = \lim_{T_k \to \infty} \frac{2}{T_k(T_k - 1)} \sum_{i<j} I_\delta(X_i^* - X_j^*), \quad l = 1, k \tag{3}$$

where $T_l = T - l + 1$ and $X_i^* = x_i$ if $l = 1$ and $X_i^* = X_i^k$ if $l = k$. Under the null hypothesis that $\{x_t\}$ are iid with a non-degenerated distribution function $F(\cdot)$, Brock et al. (1987) show that

$$C_k(\delta, T) \to [C_1(\delta)]^k \text{ with probability 1, as } T \to \infty$$

for any fixed k and δ. Furthermore, the statistic $\sqrt{T\{C_k(\delta, T) - [C_1(\delta, T)]^k\}}$ is asymptotically distributed as normal with mean zero and variance

$$\sigma_k^2(\delta) = 4\left(N^k + 2\sum_{j=1}^{k-1} N^{k-j}C^{2j} + (k-1)^2 C^{2k} - k^2 N C^{2k-2}\right), \tag{4}$$

where $C = \int \{F(z+\delta) - F(z-\delta)\} \mathrm{d}F(z)$ and $N = \int \{F(z+\delta) - F(z-\delta)\}^2$ $\mathrm{d}F(z)$.

It is to be noted that $C_1(\delta, T)$ is a consistent estimate of C and N can be consistently estimated by

$$N(\delta, T) = \frac{6}{T_k(T_k - 1)(T_k - 2)} \sum_{t<s<u} I_\delta(x_t, x_s) I_\delta(x_s, x_u) \tag{5}$$

The BDS test statistic is then defined as

$$D_k(\delta, T) = \sqrt{T\{C_k(\delta, T) - [C_1(\delta, T)]^k\}/\sigma_k(\delta, T)}, \tag{6}$$

where $\sigma_k(\delta, T)$ is obtained from $\sigma_k(\delta)$ when C and N are replaced by $C_1(\delta, T)$ and $N(\delta, T)$, respectively.

The BDS test has been performed on the daily closing values of the market indices as well as on their log returns, for embedding dimensions 2 and 3 and epsilon (eps) values of 1, 2, 3 and 4 times the observed standard deviation of the time series.

The table summarising the test results for BDS (Table 2) test of all the four data series is presented in the appendix. It can be observed from Table 2 in the Appendix that for all the values of threshold (eps) and using embedding dimension of 2 and 3, in case of all the time series the null hypothesis of iid is rejected, as indicated by the near zero *p*-value. Thus, according to the BDS test, all the fourteen time series show the presence of non-linearity.

However, the BDS test has been found to be dependent on the linear filter used. Thus there is always the concern that the rejection of the null by the BDS test could be due to the possibility of imperfect pre-whitening (Brock et al. 1991). Some of the test's sensitivity to non-linearity could be a result of remaining linear dynamics in the data.

3.3 Hinich Bi-Spectrum Test

Hinich bi-spectrum test (Hinich 1982) involves estimating the bi-spectrum of the observed time series (De Grauwe et al. 1993; Abhyankar et al. 1995; Brooks 1996; Vilasuso and Cunningham 1996). Unlike the BDS test, this test provides a direct test for a non-linear generating mechanism, irrespective of any linear serial dependencies that might be present. Thus, pre-whitening is not necessary in using the Hinich approach. Even if pre-whitening is done anyway, the adequacy of the pre-whitening is irrelevant to the validity of the test. Ashley et al. (1986) presented an equivalence theorem to prove that the Hinich linearity test statistic is invariant to linear filtering of the data, even if the filter is estimated. Thus, the linearity test can be applied to the original returns series, or to the residuals of a linear model with no loss of power.

Hinich (1982) laid out a statistical test for determining whether an observed stationary time series (y_t) is linear. The Hinich bi-spectrum test involves estimating the bi-spectrum of the observed time series, which is the double Fourier transform of the third-order cumulant function. If the process generating rates of return is linear with independent innovations, then the skewness of the bi-spectrum will be constant. If the test rejects constant skewness, a non-linear process is implied. As a consequence, the test can distinguish between white noise and purely random noise.

Let y_t denote a third-order stationary time series, where the time unit, t, is an integer.

The third-order cumulant function of y_t is defined to be $C_{yyy}(m,n) = E(y_{t+m}y_{t+n}y_t)$ for each (m,n) when $E[y_t] = 0$, in which $n < m$ and $m = 0, 1, 2, \ldots$.

Since third-order cumulants are difficult to interpret, and their estimates are even difficult to fathom, the double Fourier transform of the third-order cumulant function (called the bi-spectrum) is calculated.

The bi-spectrum at frequency pair (f_1,f_2) is the double Fourier transform of $C_{yyy}(m,n)$:

$$B_y(f_1,f_2) = \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} C_{yyy}(m,n)\exp[-i2\pi(f_1 m + f_2 n)] \tag{7}$$

assuming that $\left|C_{yyy}(m,n)\right|$ is summable. The symmetries of $C_{yyy}(m,n)$ translate into symmetries of $B_y(f_1,f_2)$ that yield a principal domain for the bi-spectrum, which is the triangular set $\Omega = \left\{0 < f_1 < \frac{1}{2}, f_1 < f_2, 2f_1 + f_2 < 1\right\}$.

Since the spectrum of y_t is $S_y(f) = \sigma_u^2\left|A_f\right|^2$

$$\psi^2(f_1,f_2) \equiv \frac{\left|B_y(f_1,f_2)\right|^2}{S_y(f_1)S_y(f_2)S_y(f_1+f_2)} = \frac{\mu_3^2}{\sigma_u^6} \tag{8}$$

for all f_1 and f_2 in Ω, where $\sum_{n=0}^{\infty} a(n)\exp(-i2\pi f_n)$.

The left hand side of Eq. (8) defines the square of the skewness function of $\{y_t, \psi(f_1,f_2)\}$. Linearity and Gaussianity of y_t are tested through the null hypotheses that $\psi(f_1,f_2)$ is constant over all frequencies and that $\psi(f_1,f_2)$ is zero over all frequencies, respectively, using the estimated bi-spectrum.

The test statistics for both hypotheses are reduced to

$$\widehat{S} = 2\sum_m\sum_n \left|\hat{y}_{m,n}\right|^2 \tag{9}$$

at the frequency pair (m, n) where

$$\hat{y}_{m,n} = \frac{\widehat{B}_y(m,n)}{\left[N/M^2\right]^{1/2}\left[\widehat{S}_y(g_m)\widehat{S}_y(g_{mn})\widehat{S}_y(g_{m+n})\right]^{1/2}}$$

Under the null hypothesis of Gaussianity, the test statistic is distributed chi-squared with 2P degree of freedom, with P being the number of squares whose centres are in the principal domain of the bi-spectrum. Hinich (1982) showed that, asymptotically, the transformation of $\widehat{S}$ is well approximated by a normal distribution with zero-mean and unit variance. Thus, the significance of the test statistics is readily determined from standard normal tables. The results (Table 3 in Appendix) reveal that the null hypothesis of linearity is rejected for all the markets. This is a strong evidence of non-linear dependence.

3.4 Keenan Test

While the non-parametric tests detect the presence of non-linearity quite strongly, another way of testing for non-linearity is through parametric methods where we test for neglected non-linearity. In Keenan (1985) test, one degree of freedom test for non-additivity has been adopted to derive a time domain statistic, as an alternative of the frequency domain statistic, e.g. bi-spectrum, for distinguishing between non-linear and linear models. The most general form of a non-linear strictly stationary process is that referred to as a Volterra expansion; this is to a linear process what a polynomial is to a linear function. Because of this similarity, an analogue of Tukey's one degree of freedom for non-additivity test is constructed as a test for linearity versus a second-order Volterra expansion.

Ramsey (1969) proposes a specification test for linear least squares regression analysis. The test is referred to as a RESET test and is readily applicable to linear AR models. We consider the linear AR(p) model

$$x_t = X'_{t-1}\varphi + a_t \tag{10}$$

where $X_{t-1} = (1, x_{t-1}, \ldots, x_{t-p})$ and $\varphi = (\varphi_0, \varphi_1, \ldots, \varphi_p)'$. The first step of the RESET test is to obtain the least squares estimate $\hat{\varphi}$ of Eq. () and we compute the fit $\hat{x}_t = \widehat{X}_{t-1}\hat{\varphi}$, the residual $a_t = x_t - \hat{x}_t$, and the sum of squared residuals $\mathrm{SSR}_0 \sum_{t=p+1}^{T} \hat{a}_t^2$, where T is the sample size. In the second step, we consider the linear regression

$$\hat{a}_t = X'_{t-1}\alpha_1 + M'_{t-1}\alpha_2 + \upsilon_t \tag{11}$$

where $M_{t-1} = \left(\hat{x}_t^2, \ldots, \hat{x}_t^{s+1}\right)'$ for some $s \geq 1$, and we compute the least squares residuals

$$\upsilon_t = \hat{a}_t - X'_{t-1}\alpha_1 - M'_{t-1}\alpha_2 \tag{12}$$

and the sum of squared residuals $\mathrm{SSR}_1 = \sum_{t=p+1}^{T} \hat{\vartheta}_t$ of the regression. The basic idea of the RESET test is that if the linear AR(p) model in Eq. (12) is adequate, then α_1 and α_2 should be zero. This can be tested by the usual F statistic of Eq. (13) given by

$$F = \frac{(\mathrm{SSR}_0 - \mathrm{SSR}_1)/g}{\mathrm{SSR}_1/(T-p-g)} \text{ with } g = s + p + 1 \tag{13}$$

which, under the linearity and normality assumption, has an F distribution with degrees of freedom g and $T - p - g$.

Keenan (1985) proposes a non-linearity test for time series that uses $\hat{x}_t^2$ only and modifies the second step of the RESET test to avoid multi-collinearity between $\hat{x}_t^2$ and X_{t-1}. Specifically, the linear regression (11) is divided into two steps. In step 2(a), one removes linear dependence of $\hat{x}_t^2$ t on X_{t-1}. by fitting the regression

$$\hat{x}_t^2 = X'_{t-1}\beta + u_t \tag{14}$$

and obtaining the residual $u_t = \hat{x}_t^2 - X'_{t-1}\beta$ In step 2(b), we consider the linear regression

$$a_t = u_t\alpha + \upsilon_t \tag{15}$$

and obtain the sum of squared residuals to test the null hypothesis $\alpha = 0$.

From Table 4 in the Appendix we find that while some of the data rejects the iid hypothesis, some do not. In case of log returns we can see that the presence of neglected non-linearity is not so significant.

3.5 *Tests Based on Neural Networks and Taylor Series Approximations*[1]

Let us consider a non-linear autoregression involving the last p lags of the variable u_t as follows

$$u_t = F\left(u_{t-1}, \ldots., u_{t-p}\right) + \varepsilon_t \tag{16}$$

Implementation of the ANN testing framework specifies that the non-linear part of $F(\cdot)$ in (16) is given by $\sum_{j=1}^{q} \beta_j \times \varphi\left(\sum_{i=1}^{p} \gamma_{ij}\hat{u}_{t-i}\right)$, where $\varphi(\lambda)$ is the logistic function, given by $[1 + \exp(-\lambda)]^{-1}$. As noted by Lee et al. (1993), this functional form can approximate any continuous function arbitrarily well. The coefficients γ_{ij} are randomly generated from a uniform distribution over $[\gamma_l, \gamma_h]$. It should be noted that using random γ_{ij} has two purposes. First, it bypasses the need for computationally expensive estimation techniques; second, and most importantly, it solves the identification problem for γ_{ij} because these parameters are not identified under the null hypothesis of linearity. For a given q, the constructed regressors $\varphi\left(\sum_{i=1}^{p} \gamma_{ij}\hat{u}_{t-i}\right), j = 1, \ldots, q$, may suffer from multi-collinearity. Following Lee

---

[1] We are immensely grateful to Prof. Timo Teräsvirta and Prof. Tae-Hwy Lee for their invaluable help in this regard.

et al. (1993), the $\tilde{q}$ in this study is taken to be the largest principle components of the constructed regressors excluding the largest one used as regressors in

$$\hat{u}_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \hat{u}_{t-i} + \sum_{j=1}^{\tilde{q}} \beta_j \tilde{\varphi}_{j,t} + \varepsilon_t, \tag{17}$$

where $\tilde{\varphi}_{j,t}$ denotes the $(j + 1)$th principal component. A standard LM test is then performed; Lee et al. (1993) suggested constructing the test statistic as TR^2, where R^2 is the uncentered squared multiple correlation coefficient of a regression of $\hat{\varepsilon}_t$ on a constant; $\hat{u}_{t-i}, i = 1, \ldots, p, \tilde{\varphi}_{j,t}, j = 1, \ldots, \tilde{q}$, where $\hat{\varepsilon}_t$ is the residual of the regression of $\hat{u}_t$ on a constant; and $\hat{u}_{t-i}, i = 1, \ldots, p$. Under the null hypothesis, this test statistic has an asymptotic $\chi^2_{\tilde{q}}$ distribution. Under the alternative hypothesis, this test is consistent, as discussed by Stinchcombe and White (1998).

An alternative two-step approach is to apply the logistic neural network test proposed by Teräsvirta et al. (1993) to the fractionally filtered series (see footnote 1). This test approximates the logistic neural network by a Taylor series expansion and tests for the significance of the additional terms when they are subsequently substituted into the model for $\hat{u}_t$. (See Blake and Kapetanios (2003) for an alternative interpretation of the logistic neural network test without long memory.) The appropriate order of terms will typically depend on the degree of non-linearity in the data. The second-order expansion is

$$\hat{u}_t = \beta_0 + \sum_{i=1}^{p} \beta_i \hat{u}_{t-i} + \sum_{i=1}^{p} \gamma_{0,i,2} \hat{u}^2_{t-i} + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma 1, i, j \hat{u}_{t-i} \hat{u}_{t-j} + \varepsilon_t, \tag{18}$$

The third-order expansion, which was recommended by Teräsvirta et al. (1993), is of the form

$$\begin{aligned} \hat{u}_t = \beta_0 + \sum_{i=1}^{p} \beta_i \hat{u}_{t-i} + \sum_{j=2}^{3} \sum_{i=1}^{p} \gamma_{0,i,2} \hat{u}^j_{t-i} + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p-1} \gamma 1, i, j \hat{u}_{t-i} \hat{u}_{t-j} \\ + \sum_{s=0}^{1} \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma 2, s, i, j \hat{u}^{2-s}_{t-i} \hat{u}^{s+1}_{t-j} + \varepsilon_t, \end{aligned} \tag{19}$$

Clearly, these are all very general approximations with a considerable number of terms and interactions. For the purpose of this study, we decided to restrict the number of parameters in the third- and fourth-order Taylor series expansions by considering only cross-products and powers of up to two lags. Given this restriction, the null hypothesis corresponding to the absence of non-linearity is equivalent to the γ coefficients being zero.

Table 4 in Appendix summarises the findings of White's Neural Network test for non-linearity in data. As is evident from the results all the series, except CSI, reject the hypothesis of linearity.

Like the White NN test, Teräsvirta Neural Network Test also shows the presence of neglected non-linearity in all the data set (Table 5 in Appendix).

4 Conclusion

As we have mentioned earlier, no one test of non-linearity dominates the others, we have used a battery of non-linearity tests to find out whether non-linearity exists in various stock markets under study. The test statistics only indicate the presence or absence of non-linearity in the time series but do not indicate the degree of non-linearity in any way. Our purpose was to find out whether non-linearity exists in all kinds of stock markets (emerging or developed). Our findings lead us to the conclusion that all the time series developed from data in the various stock markets of both the developed and emerging economies exhibit significant non-linearity for the period selected under study. While non-linear behaviour in stock markets has been evidenced in earlier studies in a segregated manner, its existence among a large diverse group of countries as revealed in this study establishes the case for further directed research along this route. It is suggested that the nature of the data generating process has to be determined first to test for non-linearity as a preliminary diagnostic tool. If the underlying data generating process for a time series is non-linear in nature, then it would be inappropriate to employ linear methods.

To progress further along empirical research on behaviour of stock markets, if non-linear behaviour of stock prices or indices is surfaced, the pattern and degree of non-linearity may be examined and compared among different countries as a future research agenda. In some studies, windows of non-linearity in stock price behaviour during a stretch of selected empirical data have been evidenced which sets the case for an in-depth analysis of the events and scenarios influencing the financial market. So recognition of non-linearity in stock indices across a wide range of countries paves the way for further analytical research in this direction.

Appendix

Table 2 The following table summarises the test results for BDS test of all the four data series

BDS test					
Developed markets					
DATA: AOI			*DATA: AOI return*		
eps[1]:	705.331		eps[1]:	0.005	
eps[2]:	1410.663		eps[2]:	0.01	
eps[3]:	2115.994		eps[3]:	0.015	
eps[4]:	2821.326		eps[4]:	0.02	
P	VALUE:		*P*	VALUE:	
eps[1]	m = 2:	0	eps[1]	m = 2:	0
eps[1]	m = 3:	0	eps[1]	m = 3:	0
eps[2]	m = 2:	0	eps[2]	m = 2:	0
eps[2]	m = 3:	0	eps[2]	m = 3:	0
eps[3]	m = 2:	0	eps[3]	m = 2:	0
eps[3]	m = 3:	0	eps[3]	m = 3:	0
eps[4]	m = 2:	0	eps[4]	m = 2:	0
eps[4]	m = 3:	0	eps[4]	m = 3:	0
DATA: DAX			*DATA: DAX returns*		
eps[1]:	959.55		eps[1]:	0.011	
eps[2]:	1919.099		eps[2]:	0.021	
eps[3]:	2878.649		eps[3]:	0.032	
eps[4]:	3838.198		eps[4]:	0.043	
P	VALUE:		*P*	VALUE:	
eps[1]	m = 2:	0	eps[1]	m = 2:	0
eps[1]	m = 3:	0	eps[1]	m = 3:	0
eps[2]	m = 2:	0	eps[2]	m = 2:	0

(continued)

Table 2 (continued)

BDS test					
Developed markets					
eps[2]	m = 3:	0	eps[2]	m = 3:	0
eps[3]	m = 2:	0	eps[3]	m = 2:	0
eps[3]	m = 3:	0	eps[3]	m = 3:	0
eps[4]	m = 2:	0	eps[4]	m = 2:	0
eps[4]	m = 3:	0	eps[4]	m = 3:	0
DATA: FCHI			*DATA: FCHI returns*		
eps[1]:	667.763		eps[1]:	0.007	
eps[2]:	1335.526		eps[2]:	0.014	
eps[3]:	2003.289		eps[3]:	0.022	
eps[4]:	2671.053		eps[4]:	0.029	
P	VALUE:		*P*	VALUE:	
eps[1]	m = 2:	0	eps[1]	m = 2:	0
eps[1]	m = 3:	0	eps[1]	m = 3:	0
eps[2]	m = 2:	0	eps[2]	m = 2:	0
eps[2]	m = 3:	0	eps[2]	m = 3:	0
eps[3]	m = 2:	0	eps[3]	m = 2:	0
eps[3]	m = 3:	0	eps[3]	m = 3:	0
eps[4]	m = 2:	0	eps[4]	m = 2:	0
eps[4]	m = 3:	0	eps[4]	m = 3:	0
DATA: FTSE			*DATA: FTSE return*		
eps[1]:	840.386		eps[1]:	0.006	
eps[2]:	1680.772		eps[2]:	0.011	
eps[3]:	2521.158		eps[3]:	0.017	
eps[4]:	3361.545		eps[4]:	0.023	

(continued)

Table 2 (continued)

BDS test					
Developed markets					
P	VALUE:		*P*	VALUE:	
eps[1]	$m = 2$:	0	eps[1]	$m = 2$:	0
eps[1]	$m = 3$:	0	eps[1]	$m = 3$:	0
eps[2]	$m = 2$:	0	eps[2]	$m = 2$:	0
eps[2]	$m = 3$:	0	eps[2]	$m = 3$:	0
eps[3]	$m = 2$:	0	eps[3]	$m = 2$:	0
eps[3]	$m = 3$:	0	eps[3]	$m = 3$:	0
eps[4]	$m = 2$:	0	eps[4]	$m = 2$:	0
eps[4]	$m = 3$:	0	eps[4]	$m = 3$:	0
DATA: HANGSENG			*DATA: HANGSENG return*		
eps[1]:	3247.651		eps[1]:	0.009	
eps[2]:	6495.301		eps[2]:	0.018	
eps[3]:	9742.952		eps[3]:	0.027	
eps[4]:	12990.6		eps[4]:	0.035	
P	VALUE:		*P*	VALUE:	
eps[1]	$m = 2$:	0	eps[1]	$m = 2$:	0
eps[1]	$m = 3$:	0	eps[1]	$m = 3$:	0
eps[2]	$m = 2$:	0	eps[2]	$m = 2$:	0
eps[2]	$m = 3$:	0	eps[2]	$m = 3$:	0
eps[3]	$m = 2$:	0	eps[3]	$m = 2$:	0
eps[3]	$m = 3$:	0	eps[3]	$m = 3$:	0
eps[4]	$m = 2$:	0	eps[4]	$m = 2$:	0
eps[4]	$m = 3$:	0	eps[4]	$m = 3$:	0

(continued)

Table 2 (continued)

BDS test					
Developed markets					
DATA: NIKKEI			*DATA: NIKKEI return*		
eps[1]:	3266.432		eps[1]:	0.007	
eps[2]:	6532.863		eps[2]:	0.015	
eps[3]:	9799.295		eps[3]:	0.022	
eps[4]:	13065.73		eps[4]:	0.029	
P	VALUE:		*P*	VALUE:	
eps[1]	m = 2:	0	eps[1]	m = 2:	0
eps[1]	m = 3:	0	eps[1]	m = 3:	0
eps[2]	m = 2:	0	eps[2]	m = 2:	0
eps[2]	m = 3:	0	eps[2]	m = 3:	0
eps[3]	m = 2:	0	eps[3]	m = 2:	0
eps[3]	m = 3:	0	eps[3]	m = 3:	0
eps[4]	m = 2:	0	eps[4]	m = 2:	0
eps[4]	m = 3:	0	eps[4]	m = 3:	0
DATA: NZX			*DATA: NZX return*		
eps[1]:	211.756		eps[1]:	0.004	
eps[2]:	423.511		eps[2]:	0.008	
eps[3]:	635.267		eps[3]:	0.011	
eps[4]:	847.023		eps[4]:	0.015	
P	VALUE:		*P*	VALUE:	
eps[1]	m = 2:	0	eps[1]	m = 2:	0
eps[1]	m = 3:	0	eps[1]	m = 3:	0
eps[2]	m = 2:	0	eps[2]	m = 2:	0

(continued)

Table 2 (continued)

BDS test					
Developed markets					
eps[2]	m = 3:	0	eps[2]	m = 3:	0
eps[3]	m = 2:	0	eps[3]	m = 2:	0
eps[3]	m = 3:	0	eps[3]	m = 3:	0
eps[4]	m = 2:	0	eps[4]	m = 2:	0
eps[4]	m = 3:	0	eps[4]	m = 3:	0
Emerging markets					
DATA: BUX			*DATA: BUX return*		
eps[1]:	554.764		eps[1]:	0.235	
eps[2]:	1109.529		eps[2]:	0.47	
eps[3]:	1664.293		eps[3]:	0.705	
eps[4]:	2219.058		eps[4]:	0.94	
P	VALUE:		*P*	VALUE:	
eps[1]	m = 2:	0	eps[1]	m = 2:	0
eps[1]	m = 3:	0	eps[1]	m = 3:	0
eps[2]	m = 2:	0	eps[2]	m = 2:	0
eps[2]	m = 3:	0	eps[2]	m = 3:	0
eps[3]	m = 2:	0	eps[3]	m = 2:	0
eps[3]	m = 3:	0	eps[3]	m = 3:	0
eps[4]	m = 2:	0	eps[4]	m = 2:	0
eps[4]	m = 3:	0	eps[4]	m = 3:	0
DATA: CSI			*DATA: CSI return*		
eps[1]:	565.929		eps[1]:	0.01	
eps[2]:	1131.857		eps[2]:	0.02	

(continued)

Table 2 (continued)

BDS test					
Developed markets					
eps[3]:	1697.786		eps[3]:	0.03	
eps[4]:	2263.715		eps[4]:	0.04	
P	VALUE:		*P*	VALUE:	
eps[1]	m = 2:	0	eps[1]	m = 2:	0
eps[1]	m = 3:	0	eps[1]	m = 3:	0
eps[2]	m = 2:	0	eps[2]	m = 2:	0
eps[2]	m = 3:	0	eps[2]	m = 3:	0
eps[3]	m = 2:	0	eps[3]	m = 2:	0
eps[3]	m = 3:	0	eps[3]	m = 3:	0
eps[4]	m = 2:	0	eps[4]	m = 2:	0
eps[4]	m = 3:	0	eps[4]	m = 3:	0
DATA: IPSA			*DATA: IPSA Retrun*		
eps[1]:	529.37		eps[1]:	0.006	
eps[2]:	1058.739		eps[2]:	0.012	
eps[3]:	1588.109		eps[3]:	0.019	
eps[4]:	2117.479		eps[4]:	0.025	
P	VALUE:		*P*	VALUE:	
eps[1]	m = 2:	0	eps[1]	m = 2:	0
eps[1]	m = 3:	0	eps[1]	m = 3:	0
eps[2]	m = 2:	0	eps[2]	m = 2:	0
eps[2]	m = 3:	0	eps[2]	m = 3:	0
eps[3]	m = 2:	0	eps[3]	m = 2:	0
eps[3]	m = 3:	0	eps[3]	m = 3:	0

(continued)

Table 2 (continued)

BDS test					
Developed markets					
eps[4]	m = 2:	0	eps[4]	m = 2:	0
eps[4]	m = 3:	0	eps[4]	m = 3:	0
DATA: KLSE			*DATA: KLSE return*		
eps[1]:	147.338		eps[1]:	0.008	
eps[2]:	294.676		eps[2]:	0.016	
eps[3]:	442.014		eps[3]:	0.023	
eps[4]:	589.352		eps[4]:	0.031	
DATA: MXX			*DATA: MXX return*		
eps[1]:	6083.843		eps[1]:	0.008	
eps[2]:	12167.69		eps[2]:	0.016	
eps[3]:	18251.53		eps[3]:	0.024	
eps[4]:	24335.37		eps[4]:	0.032	
STATISTIC:			STATISTIC:		
P	VALUE:		*P*	VALUE:	
eps[1]	m = 2:	0	eps[1]	m = 2:	0
eps[1]	m = 3:	0	eps[1]	m = 3:	0
eps[2]	m = 2:	0	eps[2]	m = 2:	0
eps[2]	m = 3:	0	eps[2]	m = 3:	0
eps[3]	m = 2:	0	eps[3]	m = 2:	0
eps[3]	m = 3:	0	eps[3]	m = 3:	0
eps[4]	m = 2:	0	eps[4]	m = 2:	0
eps[4]	m = 3:	0	eps[4]	m = 3:	0

(continued)

Table 2 (continued)

BDS test					
Developed markets					
DATA: NIFTY			*DATA: NIFTY return*		
eps[1]:	863.052		eps[1]:	0.008	
eps[2]:	1726.104		eps[2]:	0.016	
eps[3]:	2589.157		eps[3]:	0.025	
eps[4]:	3452.209		eps[4]:	0.033	
P	VALUE:		*P*	VALUE:	
eps[1]	m = 2:	0	eps[1]	m = 2:	0
eps[1]	m = 3:	0	eps[1]	m = 3:	0
eps[2]	m = 2:	0	eps[2]	m = 2:	0
eps[2]	m = 3:	0	eps[2]	m = 3:	0
eps[3]	m = 2:	0	eps[3]	m = 2:	0
eps[3]	m = 3:	0	eps[3]	m = 3:	0
eps[4]	m = 2:	0	eps[4]	m = 2:	0
eps[4]	m = 3:	0	eps[4]	m = 3:	0
DATA: TWII			*DATA: TWII return*		
eps[1]:	719.163		eps[1]:	0.008	
eps[2]:	1438.326		eps[2]:	0.016	
eps[3]:	2157.489		eps[3]:	0.024	
eps[4]:	2876.653		eps[4]:	0.032	
eps[4]	m = 3:	206.6074	eps[4]	m = 3:	11.2088
P	VALUE:		*P*	VALUE:	
eps[1]	m = 2:	0	eps[1]	m = 2:	0
eps[1]	m = 3:	0	eps[1]	m = 3:	0

(continued)

Table 2 (continued)

BDS test					
Developed markets					
eps[2]	m = 2:	0	eps[2]	m = 2:	0
eps[2]	m = 3:	0	eps[2]	m = 3:	0
eps[3]	m = 2:	0	eps[3]	m = 2:	0
eps[3]	m = 3:	0	eps[3]	m = 3:	0
eps[4]	m = 2:	0	eps[4]	m = 2:	0
eps[4]	m = 3:	0	eps[4]	m = 3:	0

To save space we have only included the p-values and excluded the absolute value of the test statistics

Table 3 Results of Hinich bi-spectrum test

Hinich bi-spectrum test for linearity and Gaussianity
Developed markets
AOI
Test statistic for Gaussianity is 1257.9681 with df = 36, Pfa = 0
AOI return
Test statistic for Gaussianity is 432.4692 with df = 36, Pfa = 0
DAX
Test statistic for Gaussianity is 70.0018 with df = 36, Pfa = 0.0006
DAX return
Test statistic for Gaussianity is 17496.9947 with df = 36, Pfa = 0
FCHI
Test statistic for Gaussianity is 59.9095 with df = 36, Pfa = 0.0075
FCHI return
Test statistic for Gaussianity is 22.5975 with df = 36, Pfa = 0.9601
FTSE
Test statistic for Gaussianity is 53.5717 with df = 36, Pfa = 0.0299
FTSE return
Test statistic for Gaussianity is 36.6538 with df = 36, Pfa = 0.04383
HangSeng
Test statistic for Gaussianity is 38.5462 with df = 36, Pfa = 0.0355
HangSeng return
Test statistic for Gaussianity is 218.279 with df = 36, Pfa = 0
NZX
Test statistic for Gaussianity is 113.5782 with df = 36, Pfa = 0
NZX return
Test statistic for Gaussianity is 34.8647 with df = 36, Pfa = 0.05226
Nikkei
Test statistic for Gaussianity is 170.3466 with df = 36, Pfa = 0
Nikkei return
Test statistic for Gaussianity is 33.7979 with df = 36, Pfa = 0.05739

(continued)

Table 3 (continued)

Hinich bi-spectrum test for linearity and Gaussianity
Developing markets
BUX
Test statistic for Gaussianity is 58.9396 with df = 36, Pfa = 0.0093
BUX return
Test statistic for Gaussianity is 42.4124 with df = 36, Pfa = 0.02138
CSI
Test statistic for Gaussianity is 216.663 with df = 36, Pfa = 0
CSI return
Test statistic for Gaussianity is 16.9736 with df = 36, Pfa = 0.0997
IPSA
Test statistic for Gaussianity is 460.7653 with df = 36, Pfa = 0
IPSA return
Test statistic for Gaussianity is 748.1778 with df = 36, Pfa = 0
KLSE
Test statistic for Gaussianity is 139.3964 with df = 36, Pfa = 0
KLSE return
Test statistic for Gaussianity is 287.7616 with df = 36, Pfa = 0
MXX
Test statistic for Gaussianity is 106.8939 with df = 36, Pfa = 0
MXX return
Test statistic for Gaussianity is 27.2389 with df = 36, Pfa = 00.8534
NIFTY
Test statistic for Gaussianity is 173.7885 with df = 36, Pfa = 0
NIFTY return
Test statistic for Gaussianity is 22.2901 with df = 36, Pfa = 0.09642
TWII
Test statistic for Gaussianity is 73.1458 with df = 36, Pfa = 0.0003
TWII return
Test statistic for Gaussianity is 16.614 with df = 36, Pfa = 0.09975

Table 4 Results of the Keenan test

Data	Test statistics	*p*-value
Developed markets		
AOI	5.348661	0.02076668
AOI return	7.846075	0.005107008
DAX	6.438327	0.01119583
DAX return	0.3648492	0.5458503
FCHI	3.646639	0.05623288
FCHI return	10.83414	0.001002609
FTSE	7.60334	0.005840813
FTSE return	5.253955	0.02192609
HangSeng	2.29346	0.1299697
HangSeng return	17.91436	2.343324e−05
Nikkei	2.048406	0.1524102
Nikkei return	4.036586	0.0445622
NZX	0.6141448	0.4333217
NZX return	34.68449	4.521386e−09
Emerging markets		
BUX	1.513421	0.2186944
BUX return	1.838422	0.175217
CSI	3.959492	0.0467637
CSI return	2.161804	0.1416588
IPSA	3.426216	0.06430161
IPSA return	8.784883	0.00306974
KLSE	0.2973768	0.5855582
KLSE return	2.743808	0.0977006
MXX	5.048076	0.02469516
MXX return	2.229881	0.1354254
NIFTY	5.716259	0.0168496
NIFTY return	5.268136	0.02176571
Taiwan	8.033133	0.00461781
Taiwan return	3.387936	0.06575497

Table 5 Results of White's neural network test

Data	*X*-squared	*p*-value
Developed markets		
AOI	3.9082	0.1417
AOI return	24.0545	5.979e−06
DAX	3.9779	0.1368
DAX return	0.4442	0.08009
FCHI	0.705	0.07029
FCHI return	4.4067	0.1104
FTSE	12.3909	0.002039
FTSE Ret	5.8865	0.0527
HangSeng	3.6211	0.01636
HangSeng return	124.2907	<2.2e−16
Nikkei	0.065	0.0968

(continued)

Table 5 (continued)

Data	*X*-squared	*p*-value
Nikkei return	52.0323	5.027e−12
NZX	0.7664	0.06817
NZX return	12.5174	0.001914
Emerging markets		
BUX	101.4424	<2.2e−16
BUX return	88.6692	<2.2e−16
CSI	1.031	0.5972
CSI return	19.3817	6.185e−05
IPSA	3.5154	0.1724
IPSA return	30.5269	2.351e−07
KLSE	0.0593	0.09708
KLSE return	329.8214	<2.2e−16
MXX	6.1714	0.0457
MXX return	45.0401	1.658e−10
NIFTY	13.3653	0.001252
NIFTY return	22.3315	1.415e−05
Taiwan	3.5535	0.01692
Taiwan return	6.8163	0.0331

Table 6 Results of Teräsvirta neural network test

Data	*X*-squared	*p*-value
Developed markets		
AOI	4.1297	0.1268
AOI return	8.7987	0.01229
DAX	6.62086	0.0365
DAX return	0.3648492	0.0456
FCHI	1.4181	0.4921
FCHI return	21.6111	2.029e−05
FTSE	12.8824	0.001595
FTSE Ret	42.1527	7.025e−10
HangSeng	5.3096	0.07031
HangSeng return	245.7963	2.2e−16
Nikkei	0.0078	0.9961
Nikkei return	99.5254	<2.2e−16
NZX	1.4634	0.4811
NZX return	18.2972	0.0001064
BUX	146.0156	<2.2e−16
Developing markets		
BUX return	141.7391	<2.2e−16
CSI	1.127	0.5692
CSI return	23.5759	7.596e−06
IPSA	3.8232	0.1478
IPSA return	28.8542	5.425e−07
KLSE	0.3646	0.8334

(continued)

Table 6 (continued)

Data	*X*-squared	*p*-value
KLSE return	447.6535	<2.2e−16
MXX	5.544	0.06254
MXX return	98.5446	<2.2e−16
NIFTY	14.0483	0.0008901
NIFTY return	9.7106	0.007787
Taiwan	4.1055	0.1284
Taiwan return	13.4813	0.001182

References

Abhyankar AH, Copeland LS, Wong W (1995) Nonlinear dynamics in real-time equity market indices: evidence from the United Kingdom. Economic Journal 105:864–880

Abhyankar AH, Copeland LS, Wong W (1997) Uncovering nonlinear structure in real time stock market indexes: the S&P 500, the DAX, the Nikkei 225, and the FTSE 100. Journal of Business and Economic Statistics 15:1–14

Afonso A, Teixeira J (1998) Nonlinear tests of weekly efficient markets: evidences from portugal. www.papers.ssrn.com/sol3/…/ssrn_ID317485_code020721500.pdf

Ashley RA, Patterson DM, Hinich MJ (1986) A diagnostic test for nonlinear serial dependence in time series fitting errors. J Time Ser Anal 7:165–178

Baek E, Brock W (1992) A general test for non-linear Granger causality: bivariate model. Working paper. Iowa State University and University of Wisconsin at Madison

Barkoulas J, Travlos N (1998) Chaos in an emerging capital market? The case of the athens stock exchange. Appl Financ Econ 8:231–243

Barnett WA, Gallant AR, Hinich MJ, Jungeilges J, Kaplan D, Jensen MJ (1997) A single-blind controlled competition among tests for nonlinearity and chaos. J Econ 82:157–192

Barnett WA, Serletis A (2000) Martingales, nonlinearity, and chaos. J Econ Dyn Control 24:703–724

Bhattacharya A, Sensarma R (2013) Non-linearities in emerging financial markets: evidence from India. IIM Kozhikode Society & Management Review 2 Feb 2013, pp. 165–175

Bierens HJ (1997) Nonparametric cointegration analysis. J Econ 77:379–404

Blake AP, Kapetanios G (2003) A radial basis function artificial neural network test for neglected nonlinearity. Econ J

Blasco N, Rio CD, Santamaria R (1997) The random walk hypothesis in the Spanish stock market: 1980–1992. J Bus Finan Account 24(5):667–683

Brock W, Dechert WD, Scheinkman J (1987) A test for independence based on the correlation dimension. SSRI working paper no. 8702, Department of Economics, University of Wisconsin, Madison, WI

Brock WA, Dechert WD, Scheinkman JA, LeBaron B (1996) A test for independence based on the correlation dimension. Econ Rev 15:197–235

Brock WA, Hsieh DA, LeBaron B (1991) Nonlinear dynamics, chaos, and instability:statistical theory and economic evidence. MIT Press, Cambridge

Brooks C (1996) Testing for non-linearity in daily Sterling exchange rates. Appl Financ Econ 6:307–317

Campbell JY, Lo AW, MacKinlay AC (1997) The econometrics of financial markets. Princeton University Press, Princeton

Chortareas GE, Kapetanios G, Shin YC (2002) Non-linear mean reversion in real exchangerates. Econ Lett 77:411–417

Cinko M (2002) Nonlinearity test for Istanbul stock exchange. http://www.cu.edu.tr/sempozyum/bi/3.htm

Dahl CM, Nielson S (2001) The random walk of stock prices: Implications of recent non parametric tests. Working Paper, Frederiksberg: Department of Economics

De Grauwe P, Dewachter H, Embrechts M (1993) Exchange rate theory: chaotic models offoreign exchange markets. Blackwell, Oxford

Dorina L, Simina U (2008) Testing efficiency of the stock market in emerging economies. www.steconomice.uoradea.ro/anale/volume/2007/v2-statistics…/24.pdf

Hassan KM, Al-Sultan WS, Al-Saleem JA (2003) Stock market efficiency in the Gulf Council Countries (GCC): The case of Kuwait stock exchange. Sci J Adm Dev 1(1):1–21

Hinich MJ (1982) Testing for Gaussianity and linearity of a stationary time series. J Time Ser Anal 3:169–176

Hinich MJ, Patterson DM (1985) Evidence of nonlinearity in daily stock returns. J Bus Econ Stat 3:69–77

Hinich MJ, Patterson DM (1989) Evidence of nonlinearity in the trade-by-trade stock marketreturn generating process. In: Barnett WA, Geweke J, Shell K (eds) Economic complexity: chaos, sunspots, bubbles and nonlinearity-international symposium in economic theory and econometrics. Cambridge University Press, Cambridge, pp 383–409

Hiremath GS, Kamaiah B (2010) Non-linear dependence in stock returns: evidence from India. J Quant Econ 8(1):69–85

Hsieh DA (1988) The statistical properties of daily foreign exchange rates: 1974–1983. J Int Econ 24:129–145

Hsieh DA (1989) Testing for non-linearity in daily foreign exchange rate changes. J Bus 62: 339–368

Hsieh DA (1991) Chaos and nonlinear dynamics: application to financial markets. J Financ 46:1839–1877

Kapetanios G, Shin YC, Snell A (2003) Testing for a unit root in the nonlinear STAR framework. Journal of Econometrics 112:359–379

Keenan DM (1985) A Tukey non-additivity-type test for time series nonlinearity. Biometrika 72: 39–44.

Lee TH, White H, Granger CWJ (1993) Testing for neglected nonlinearity in time series models. J Econ 56:269–290

Lim K-P, Brooks R (2009) Are Chinese markets efficient? Further evidence from a batter of nonlinearity tests. Appl Financ Econ 19(2):147–155

Lima JF (1998) Nonlinearities and non-stationarities in stock returns. J Bus Econ Stat 16(2): 227–236

Opong KK, Mulholland G, Fox AF, Farahmand K (1999) Thebehavior of some UK equity indices: an application of Hurst and BDS tests. J Empir Financ 6:267–282

Pagan A (1996) The econometrics of financial markets. J Empir Financ 3:15–102

Pesaran MH, Potter SM (1993) Nonlinear dynamics, chaos and econometrics: an introduction. In: Pesaran MH, Potter SM (eds) Nonlinear dynamics, chaos and econometrics. John Wiley & Sons, New York, pp. vii–xiii

Ramsey JB (1969) Tests for specification errors in classical linear least squares regression analysis. J Roy Stat Soc B 31:350–371

Sarno L (2000) Real exchange rate behaviour in high inflation countries: empirical evidence from Turkey, 1980–1997. Appl Econ Lett 7:285–291

Sarno L (2001) Thebehavior of US public debt: a nonlinear perspective. Econ Lett 74:119–125

Scheicher M (1996) Evidence for a small stock exchange. Working paper, Department of Economics, BWZ University of Vienna, Vienna, no. 9607

Scheinkman J, LeBaron B (1989) Nonlinear dynamics and stock returns. J Bus 62:311–337

Seddighi HR, Nian W (2004) The Chinese stock exchange market: Operations and efficiency. Appl Financ Econ 14:785–797

Sewell SR, Lee SI, Pan MS (1993) Nonlinearities in emerging Foreign capital markets. J Bus Financ Account 20:237–248

Sharma A, Panagiotidis T (2005) An analysis of exports and growth in India: cointegration and causality evidence (1971–2001). Rev Dev Econ 9(2):232–248

Steurer E (1995) Nonlinear modelling of the DEM/USD exchange rate. In: Refenes AP (ed) Neural networks in the capital markets. Wiley, New York, pp. 199–211

Stinchcombe MB, White H (1998) Consistent specification testing with nuisance parameters present only under the alternative. Econ Theor 14:295–325

Taylor MP, Peel DA (1997) Nonlinearities in real exchange rate adjustment during the recent float: empirical evidence and Monte Carlo analysis. Institute of Economics and Statistics, University of Oxford, Discussion Paper

Taylor MP, Peel DA (2000) Non-linear adjustment, long-run equilibrium and exchange rate fundamentals. J Int Money Financ 19:33–53

Teräsvirta T, Fu-Lin C, Granger CWJ (1993) Power of the neural network test. J Time Ser Anal 14(2):209–220

Tsay R (2005) Analysis of financial time series. Wiley, New Jersey

Vilasuso J, Cunningham S (1996) Tests for nonlinearity in EMS exchange rates. Stud Nonlinear Dyn Econ 155–168

Willey T (1992) Testing for nonlinear dependence in daily stock indices. J Econ Bus 44(1):63–76

Part VI
Issues Related to Foreign Exchange Market

Foreign Exchange Markets, Intervention, and Exchange Rate Regimes

Ashima Goyal

Abstract While macroeconomic fundamentals determine the exchange rate at long horizons, there are substantial and persistent deviations from these fundamentals. The market microstructure within which they operate, macroeconomic fundamentals, and policies all affect foreign exchange (FX) markets. The chapter describes the institutional features of these markets, with special emphasis on the process of liberalization and deepening in Indian FX markets, in the context of global integration. Since the mechanics of FX trading affect exchange rates, they have implications for the appropriate exchange rate regime. First, bounds on the volatility of the exchange rate can lower noise trading in FX markets, decrease variance, improve fundamentals, and give more monetary policy autonomy. Second, the speculative demand curve is well behaved under strategic interaction between differentially informed speculators and the Central Bank (CB) when there is greater uncertainty about fundamentals as in emerging markets. So, a diffuse target and strategic revelation of selected information can be expected to be effective. Analysis of Indian experience confirms these research results. CB actions, including intervention and signaling, have major effects.

Keywords Foreign exchange markets · Intervention · Information · Exchange rate bounds

JEL Codes F41 · G15

1 Introduction

Rapid technological and regulatory changes are altering foreign exchange (FX) markets profoundly, although special features of these markets are likely to survive. Participant profile and behavior is also changing, as technology makes it

---

A. Goyal (✉)
Indira Gandhi Institute of Development Research, Gen. Vaidya Marg, Santosh Nagar, Goregaon (E), Mumbai 400065, India
e-mail: ashima@igidr.ac.in

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_23

easier for small parties to access markets anytime and anywhere. This chapter reviews these changes, internationally and then domestically. Reforms in Indian markets offer a good case study of the tensions between regulators, markets and technology. The chapter also draws out the implications for effective central bank (CB) interventions and for exchange rate policy.

The current consensus from many empirical studies is while macro fundamentals determine the exchange rate at long horizons, there are substantial and persistent deviations largely unexplained by these fundamentals.

A literature on market microstructure of FX markets grew rapidly after seminal work by Lyons (2001). This suggests the mechanics of FX trading has important short-run effects on exchange rates. This is a departure from the traditional modeling strategy of treating foreign exchange rates as a macroeconomic relative price. It also implies it is not only public information which is relevant.

Survey data shows that there is considerable heterogeneity in agents' expectations of the future exchange rate. Therefore, private information, transmission of this information, and relation between information flows plays a vital role. Proprietary information is contained in order flows, which is the net of buyer-initiated and seller-initiated orders. Although this is a variant of net demand, it is not necessarily equal to zero in equilibrium. Each bank will have knowledge only of its own order flow, which is used to update subjective estimates of the underlying value of the currency.

Order flow is a more precise proxy for expected future fundamentals, since it presents a willingness to back one's beliefs with real money. In specifications that include macro fundamentals and order flow variables, Lyons (2001) finds order flow to be a significant determinant of the exchange rate. It performs better than both standard macroeconomic variables, and random walk forecasts.

It follows foreign exchange markets are influenced by a combination of macroeconomic and microeconomic variables. The market microstructure within which they operate, macroeconomic fundamentals, and policies affect the decisions micro market agents make. The Central Bank (CB) is a special kind of agent with special powers and information. So its actions, including intervention and signaling, also have major effects. The Indian experience confirms this.

Other participants in FX markets include banks, non-bank financial companies (NBFCs), merchants and merchant-brokers. Large banks play the role of market makers, accepting both buy and sell quotes. Merchant transactions were originally restricted to trade and other retail transactions involving foreign currency, but now generate many kinds of transactions due to risk management activities. Large and sophisticated corporate treasuries now have multiple FX operations.

The structure of the chapter is as follows: Sect. 2 discusses some institutional features of foreign exchange markets, with a special sub-section and boxes on the process of liberalization and deepening in Indian FX markets. Section. 3 draws out implications of varying trader information for exchange rate policy. Section. 4 concludes. An Appendix gives some derivations for the model used in Sect. 3.

2 Foreign Exchange Market: Institutional Features

Compared to other financial markets, FX markets have unique features.[1] We briefly describe their structure, composition, effects of change in technology and in regulations, then draw out implications for their functioning.

Structure: They are the most liquid markets. Daily market turnover was $5.4 trillion in 2013 (BIS 2013). But only about 5 % of the very large turnover is actually due to customer trade. Decentralized large volume markets with many physically separated market makers, interact through the telephone or private networks, not in a centralized market like a stock market. Decentralization makes FX markets fragmented and less transparent. There is no publicly announced price and no law requiring disclosure of trades. Each broker or market maker only knows own order flows, with no incentive to share the information. Brokers normally accumulate a subset of market makers' limit orders, and quote the best buy and sell order from a 'book' they keep of such limit orders. A limit order is an offer to either buy or sell a certain quantity of a currency at a certain bilateral exchange rate.

In a stock market clearing-house, each party trades with it—doubling the number of transactions. This is known as "novation." The identity of the counterparty does not matter since the clearing-house warranties the trade covering its risk through margin payments and deposits. But in an FX market there are many market makers. Market spreads can vary to cover the cost of market-making including counterparty risk.[2] Even if the net position is close to zero, credit limits get filled up, unlike in a clearing-house, where only the net position is required for settlement.

Banks that are reporting dealers have to be the market makers in a decentralized market since brokers cannot assess credit worthiness. Identity may be known only after the deal. The direct FX market is double auction and open-bid. That is, two-way prices on both bid and ask are announced to all agents in the market. The brokered market is single auction and limit order. That is, prices are specified only to buy or sell but not both. They are known only to the broker and the party making the offer.

Counterparties, instruments, currencies: Participants are heterogeneous with diverse information sets and reaction speeds, so that profit opportunities can persist for informed traders. Central banks have a special position. Although the interbank market continued in 2013 to account for the majority of transactions (63 %) this share decreased since the nineties because of the rise of other financial institutions including groups such as institutional investors, hedge funds (22 %) and small non-reporting banks (24 %). Sudden shifts in positioning by large hedge funds that have the fastest reaction speeds and operate with high leverage can magnify shocks

---

[1]This section is largely based on material in Lyons (2001), Sarno and Taylor (2002), Sager and Taylor (2006), BIS triennial CB surveys and media reports.

[2]Counterparties made large losses as currency volatility spiked after Lehman fell in 2008. The risk to market makers inventory caused spreads on quotes to increase from 4 to 16 pips. Trade froze for some transactions.

to FX markets. They implement currency programs to secure a notional capital value that may be a benchmark risk-free rate.

As the corporate treasury and direct mobile trading FX market grew, traditional brokers were by-passed. But prime brokerage relationships with their clients dominated dealers' trade (16 %) with only 3.5 % driven by trade with retail customers. Corporate treasuries became sophisticated. Customers changed from passive price takers with emphasis on financing and other banking services relating to foreign trade, foreign investors, corporates availing foreign borrowing or involved in mergers and acquisitions, etc. But the share of these nonfinancial customers in trading fell overall to 9 % in 2013.

In April 2013 the US dollar as the dominant vehicle currency was on one side of 87 % of all trades. But renminbi became the 9th most traded currency as turnover grew rapidly to $120 billion. The financial centers UK (with 41 %), US, Singapore, and Japan intermediated 71 % of FX trading. Post GFC bank closures concentrated trading increasingly in the large banks. The ten most active global traders accounted for 77 % of trading volume, of which the top three, Deutsche, Citi and Barclays, were at 40 %, according to the 2012 Euromoney FX survey.

OTC turnover at about 95 % share continued to far exceed turnover on exchanges. In 2013 OTC FX swaps were the most actively traded at 42 %, but forwards and options slowed the most rapid growth.

Technology: Although new technologies are causing some change, the majority of transactions continue to be bilateral, occurring in opaque markets without a physical market place. Even so, electronic dealing and brokering systems, are giving some amount of virtual centralization. Electronic Broking System (EBS) or Reuters D3000, established in 1993, accounted for 85 % of interbank trading by the 2000s.

Electronic systems allow netting, lower settlement and counterparty risk, and have operational benefits such as reducing human error. They provide ex-ante anonymous limit order bid-ask pricing to dealers and have driven a large increase in liquidity and reduction in transaction costs. CLS, the continuous link settlement system used by the majority of the FX market, settles payment instructions relating to underlying FX transactions in 17 major currencies, reducing settlement risk. The share of inter-dealer trade (39 %) fell as increasing concentration allowed dealers to match customer trade on their own books, and investment in IT infrastructure for warehousing risk reduced the need to offload inventory in the inter-dealer market.

Although voice trading dominates in customer trades, electronic portals are being introduced here also. Electronic crossing networks (ECNs) that aggregate liquidity pools received a fillip from the global financial crisis (GFC). As markets froze, they were successful in finding liquidity where trade could occur without a large impact on price. But electronic systems do not increase the transparency of the FX market since system governing boards treat electronic order flow as strictly confidential. Therefore, information on order flow remains divided.

Regulation: It remains to be seen if the regulatory push towards greater transparency after the GFC, which is shifting more over-the-counter (OTC) trade to exchanges, causes a fundamental change. But even the US Dodd-Frank Act that sought to prevent banks' proprietary trading, has given exceptions for the spot FX market, thus accepting that the FX market is different. Higher capital requirements and tighter regulations are, however, reducing banks' participation in all markets.

Despite magnified activity, currency markets remained largely stable during the financial crisis of 2008, partly because risk management procedures had been improved after earlier crises. Banks imposed position limits for individual traders, and risk capital made available was a function of past performance. Incentives to take risk were reduced because losses reduced traders' risk capital while profits were shared with the bank (Geithner 2004).

But regulation has to continue to evolve in response to new types of malpractices. Since banks often act as principal to a trade, they buy at the moment the client sells. This conflict of interest gives them an incentive to move rates against their customers. Such behavior is difficult in a transparent competitive market since customers getting a poor rate would move elsewhere. But FX markets are not transparent and collusion further removes such protection. In 2014, traders were caught fixing benchmark rates to suit their own positions. Employees exchanged confidential client information with rival firms in order to trigger orders against their own customers thus distorting the market. In November regulators from the UK, Switzerland, and the US imposed record fines of US $4.3 billion on six large banks whose weak controls allowed these malpractices. More fines followed.

Decentralized currency trading with huge volumes scattered across numerous platforms makes it difficult to monitor and identify dubious trades. Solutions being considered include extending the period during which the daily fixed rate is established to make it harder to manipulate. Big data is being used in creative ways to flag unusual activities. Fines reduce the financial incentive to cheat. They also create pressure on management from bank shareholders. Values set by the top management are also important. Traders caught are normally either dismissed or lose their bonus. There are suggestions for higher penalties include the risk of a jail sentence. Solutions therefore range from better monitoring and incentives to the role of values.

Functioning: More transparency could also be a possible solution. There is some rise in this. More trading on exchanges creates price benchmarks. But is it possible to change the decentralized largely OTC structure of the market with its huge trading volumes? Or does it serve some purpose? Large temporary inventory imbalances generate 'hot potato' trading as dealers iterate towards their optimal portfolios although the share of such inter-dealer trade is falling, it remains very large. Market makers and dealers do not want to carry inventory overnight—which carries inventory risk—therefore they quote ask (buy) and bid (sell) spreads such as to get rid of stocks in the day.

Trades are initiated based on macro data and differential order flow information, with the aim of rebalancing portfolios. The information in the order flow sustains trade. The transactions are not all speculative, or profit seeking. Although this market structure raises the number of transactions, it is less prone to crashes. A centralized system with too many informed traders would crash as liquidity dried up due to homogeneous views, especially given the few prices quoted.[3] Compared to the large number of quotes in a stock market, prices in an FX market refer only to a few currency pairs, making the FX market more susceptible to herding and explaining its differential structure.

2.1 Indian FX Markets

Indian FX markets offer an interesting case study of the process of market development. Intra-day trade was first permitted for banks in 1978, but the market really grew after liberalization,[4] as The Sodhani Committee's (RBI, 1995) comprehensive blueprint for reform was followed. The Tarapore Committee (2006) also made several recommendations for these markets. Despite major changes in the expansion of turnover and of instruments available for hedging they were still far behind international markets. The advent of electronic trading and communication platforms, reduced transaction costs and risks, and the profile of customers as capital flows became the prime mover of exchange rates. Rising exchange rate volatility, with a more open capital account, increased the necessity for hedging FX risks.

The average daily turnover in Indian OTC FX markets, which was about US $2.0 billion in 1998, grew to US $38 billion in 2007.[5] Growth slowed after the GFC, but even so by April 2010 the daily domestic OTC market turnover was $27 billion and the futures market about $10 billion. So unlike the global average of 4 % the Indian exchange traded market was about 30 % of the domestic market (Mecklai 2010b). BIS measures only OTC market turnover. The interbank to merchant turnover ratio halved from 5.2 during 1997–98 to 2.3 during 2007–08 reflecting the growing participation in the merchant segment of the foreign exchange market. The spot market remained the most important FX market segment accounting for above 50 % of the total turnover. Its share also declined marginally due to a pick-up in the turnover in the derivative segment. Even so, Indian derivative trading remained a small fraction of that in other developing countries such as Mexico or South Korea. Short-term instruments with maturities of less than 1 year dominated, and activity was concentrated among a few nonpublic sector banks (IMF 2008).

---

[3]For example Indian FX futures markets grew rapidly, after they were established, but were still thin. If a large party came in on the buy side the sell side would dry up in anticipation of a price rise.

[4]This section, unless explicitly mentioned, updates information in Goyal et al. (2009) and Goyal (2015).

[5]BIS (2007) notes this was the fastest rate of growth amongst all world FX markets, although the 72 % rate of growth of world FX market activity between 2004 and 2007 was also the fastest. In the next 3 years growth was 19 % but rose to 35 % over 2010–13.

Table 1 Comparison of Indian and Australian FX markets

USD billion		Australia			India		
		2001	2007	2013	2001	2007	2013
Daily FX turnover	Amount	54 (54)	176 (220)	182 (462)	3 (3)	38 (24)	31 (53)
	%	3.2	4.1	2.7	0.2	0.9	0.5
Merchandise trade, daily average		0.02	1.1	1.7	0.4	1.5	2.8
FX inflow, daily average		0.02	0.07	0.14	0.02	0.26	0.16

Note (1) Foreign inflows are measured as the current account deficit plus reserve gains. (2) Merchandise trade is calculated as exports plus imports of goods and services (absolute values) (3) Domestic FX turnover is on net-gross basis, (that is adjusted for local inter-dealer double counting by subtracting half of the turnover with reporting local dealers). It includes spot, outright forwards and swap transactions. Global INR turnover is given on a net-net basis in brackets. This adjusts for local and cross-border inter-dealer double counting by subtracting half of the turnover with reporting dealers abroad. BIS (2013) warn turnover for years prior to 2013 may be under-reported, especially for EMs
Source FX turnover calculated from the Bank for International Settlements, various years, for example, (BIS 2007, Table E16, pp. 82, http://www.bis.org/publ/rpfxf07a.pdf), the International Financial Statistics (IMF, various years)

Box 1: Deepening of Indian FX Markets Table 1 shows the rapid deepening of Indian FX markets, the rise in trade and inflows which are dwarfed by the large turnover which itself still remains small even in comparison a middle level developing country like Australia. Table 2 shows the shares by types of agents and instruments, with domestic market data from the RBI, and global INR trade (row 8 onwards) from the BIS. With deepening there is a sharp fall in the share of CB transactions and some rise in derivative use although regulators restraints slowed these after the Euro-debt crisis of 2011, but only domestically. Cross border transactions also rose. The share of derivatives is much higher in global INR trade (row 15 Table 2) pointing to a large offshore market. Daily global net-net INR turnover is also higher than domestic FX market turnover (Table 1).

The percentage of intervention to interbank turnover fell from 13.4 in 2001–02 to 0.9 in 2006–07, but it was still large compared to mature economies. The Bank of Japan intervened successfully in 2011 even with a percentage of 0.2. This is the annual intervention percentage. The CB share can be much higher for daily intervention, which tends to be concentrated on a few days. Since the interbank market remains a large size of the total, the interbank share is not much higher than the percentage of CB intervention to total turnover. CB intervention, however, affected only domestic markets.

Even so, the derivative segment of the FX market also evolved. Cross-currency derivatives with the rupee as one leg were introduced, with some restrictions, in April 1997. Rupee-foreign exchange options were allowed in July 2003. Exchange-traded currency futures were started in 2008.[6] The most widely used derivative instruments were the forwards and foreign exchange swaps (rupee–dollar). As elsewhere, FX transactions were mostly OTC structured by banks. But

[6]In the absence of full rupee convertibility, a future contract could not result in the delivery of foreign currency. It was netted out in rupees, reducing its usefulness for hedging.

Table 2 Aspects of the Indian FX Market

	US $ billion FCY/INR[a]	2001–02	2006–07	2012–13
1	Total domestic spot turnover (sales + purchases)	446.1	1861.4	4525.2
2	Total CB intervention (sales + purchases)	38.6	26.8	29.9
3	2 as % of 1	8.7	1.5	0.7
4	Share of 1 due to interbank (%)	64.5	66.3	73.4
5	Share of 1 due to merchant (%)	35.6	33.7	26.6
6	Total forward as % of total spot	22.5	23.6	24.4
7	Total swap as % of total spot[b]	147.4	77.2	75.1
8	Global total INR spot (for April) (OTC)[c]	1.2	9.0	15.2
9	Share due to RDs (from CB survey) (%)	51.2	63.1	45.2
10	Share due to other financial insts. (%)	9.8	18.4	38.1
11	Share of non-financial insts. (%)	39.1	18.5	16.7
12	Share in total spot of local transactions(%)	94.1	77.2	66.9
13	Share in total spot of cross border tran. (%)	5.9	22.8	33.1
14	Total domestic FX derivatives as % of total spot (net-gross)	116.8	137.5	102.1
15	Total global INR FX derivatives as % of total spot (net-net)	110.9	134.5	246.5

Note Items (1) to (7) were calculated from RBI bulletins. The data was collected for all the months in the given years and summed up. Each year is taken from April to March. (8) to (15) are available in the Central Bank (CB) Surveys (BIS) and refer to net-net daily averages added up across different participants for April 2001, 2007, and 2013, respectively. Items (9) to (13) and (15) are as percentage to (8), (14) is a percentage of spot in net-gross terms; FCY Foreign currency; INR Indian rupees; RDs Reporting dealers

[a]All transactions involve exposure to more than one currency

[b]Excluding "tomorrow/next day" transactions

[c]A swap is considered to be a single transaction in that the two legs are not counted separately. Including "tomorrow/next day" transactions

there was user demand for liquid and transparent exchange traded hedging products, which are easier to regulate.

The non-deliverable forward (NDF) OTC market was growing because of the large capital and trade flows, restrictions on FIIs ability to hedge in domestic markets, and larger spread between forward, futures and NDF markets. It began with diamond traders using it for arbitrage. In 2008, the Indian forward market was fairly liquid up to 1 year. The price movement in the near-term bucket reflected rupee liquidity in the interbank market and overnight interest rates but the 6-month and 1-year rates were determined also by expected future liquidity. Importers and exporters also influenced the forward markets. Forward rates in a particular segment could differ from other segments due to the excess supply/demand from importers/exporters in that segment.

The Clearing Corporation of India Ltd. (CCIL) set up by the Reserve Bank of India (RBI) in 2001 settled 90–95 % of interbank rupee–dollar transactions. Foreign exchange trades were settled through multilateral netting thus saving transaction cost. All spot, cash, tom transactions, and forward trades were guaranteed for settlement from the trade date reducing foreign exchange settlement and counterparty risk. A transparent FX dealing system, FX-Clear, of the CCIL launched in August

Table 3 Policy measures over 2010–2014 and effect on rupee (+depre, −appre)

Date	Change in INR/USD (week before)	Change in INR/USD (week after)	Policy action
28 December 2010	−0.21	0.03	RBI issues guidelines for OTC FX derivatives and overseas hedging
1 February 2011	0.11	−0.48	Derivatives guidelines applied
15 September 2011	1.82	1.83	Exchange earners foreign currency account and residents foreign currency accounts-liberalization
15 November 2011	1.19	1.54	Increase in ceiling rate on banks' export credit in foreign currency by 150 basis points
5 December 2011	−0.77	2.18	Speech reinforcing RBI's hands-off policy
15 December 2011	2.79	−1.51	Bank net open position limits (NOPL) reduced 75 %; Free cancelation and rebooking of FX forward contracts disallowed
21 May 2012	1.04	0.90	Netting of positions in currency futures/options with OTC positions disallowed; position limits of banks for currency futures and options reduced
11 September 2012	0.07	−1.18	ECB policy eased
13 May 2013	0.96	0.12	RBI restricts banks' gold imports
22 May 2013	0.89	0.58	Bernanke says Fed may taper QE
20 June 2013	1.43	−0.001	Foreign banks open positions in USD/INR reduced to almost zero
9 July 2013	0.93	−0.71	Any proprietary activity by banks in currency futures banned
10 July 2013	0.72	−0.42	Public sector oil companies directed to buy FX only from one bank (SBI)
23 July 2013	−0.36	1.43	Monetary tightening measures started from July 9; reduced LAF limit to 0.5 % of a bank's own NDTL; banks to maintain a daily minimum CRR balance of 99 %; MSF rate raised to 10.25 and CMR moved up to it from repo of 7.25
28 August 2013	4.63	−2.32	FX swap window for oil companies (closed end-November)

(continued)

Table 3 (continued)

Date	Change in INR/USD (week before)	Change in INR/USD (week after)	Policy action
4 September 2013	1.36	−3.24	Window for the banks to swap the fresh FCNR(B) deposits with RBI and increase in Banks' overseas borrowing limit with option of swap with RBI
18 September 2013	−1.07	−0.92	Fed refrains from QE taper, keeps bond buying at US $85 billion
11 November 2013	1.89	−0.74	Participation by SEBI registered FIIs, QFIs long-term investors in credit enhanced bonds
21 November 2013	−0.63	−0.80	Eased bank's use of swaps in negotiation of loans from international/multilateral financial institutions.
28 January 2014	1.36	−0.54	Fed reduces QE
3 September 2014	0.12	0.37	Relaxation of External Commercial Borrowings (ECB) limits
28 October 2014	−0.13	0.17	Fed ends QE
28 November 2014	−0.13	−0.05	Oil prices plummet as OPEC refuses to cut production

Source Updated from Goyal (2015)

2003, decreased settlement risk and gave netting and operational benefits. It facilitated interbank trade through order matching and negotiation mode. Reuters platform was also available. Swaps and options were essentially inter-bank transactions, and accounted for about 50 % of CCIL trade settlement (IMF 2008).

The Reserve Bank moved gradually to eliminate restrictions on FX markets. Historically, the availability of hedging tools against foreign exchange risk was limited to entities with direct underlying foreign exchange exposures. However, with a larger set of economic agents exposed to foreign exchange risk there was a shift to the concept of "economic exposure," that is, the effect of exchange rates on a firm's value. There were gradual steps to give greater flexibility to corporates for managing their exposures. For example, it was proposed to permit agents to book forward contracts without production of underlying documents up to an annual limit of US $100,000, which could be freely canceled and rebooked. Cancelation and rebooking of forward contracts and swaps in India were regulated to reduce rupee volatility. There were moves to allow banks to fix their own net open position limits (NOPL) and AGL limits based on their risk appetite and ability to manage exposure, with adequate prudential regulation and supervision to cover systemic risk and prevent excessive leverage. By 2011, while banks boards set the NOPLs they had to be approved by the RBI.

Box 2: The Process of Regulatory Change in India FX market regulations followed a dynamic process driven by regulatory objectives of market development with stability, demands from and requirements of markets. Some examples of this dialectic are given below, over 2002–2013, a period with major changes in Indian FX markets.

Since 2002, persons resident in India were allowed to enter into forward contracts on the basis of underlying exposures. Further, exporters and importers were allowed to book forward contracts on the basis of declaration of exposures and based on past performances, subject to specified conditions. Permissions were slowly expanded, with the aim of enabling hedging through the reversal of a real transaction.

The annual Policy Statement for the year 2007–2008 (paras 142) provided greater flexibility to the Small and Medium Enterprises (SME) sector and resident individuals, further liberalization of the scope and range of forward contracts, to facilitate such entities to hedge their foreign currency exposures on a dynamic basis. There was a warning that authorized dealer (AD) Category—I bank should carry out due diligence regarding "user appropriateness" and "suitability" of the forward contracts to the SME customers.

NRIs could now book forward contracts without production of underlying documents up to a limit of US $100,000, based on self-declaration. These contracts would normally be on a deliverable basis. However, in case of mismatches in cash flows or other exigencies, the contracts booked under this facility could be canceled and rebooked. The notional value of the outstanding contracts was not to exceed US $100,000 at any time. Further, the contracts were permitted for tenors of up to 1 year only.

Source: RBI/2007–2008/, A. P. (DIR Series) Circular No. October 10, 2007

In an interview conducted in Sept. 2007, Mr. Bhaskar Panda—senior vice-president and regional head—treasury advisory group—HDFC Bank, assessed the changes and advocated further reform as follows:

> Customized options have mostly evolved over the past 4-5 years after RBI liberalized its norms. Earlier, a corporate could hedge its risk only for 3 years, today they can hedge it for up to 10 years. But the value of the hedge is capped up to the basis of last year's turnover. Banks' want this regulation be altered to allow booking of forward contracts based on projected performances. Banks mostly trade on Reuters terminal, CCIL and voice brokers. Technology has made a big difference to the level of FX dealing and has helped significantly to increase volumes.

CCIL was guaranteeing forward trades from the date they entered the spot window. But huge outstanding FX exposure and capital requirements still remained. Member banks wanted CCIL to extend guarantee to these trades from trade date itself. This would imply reduction in bilateral exposure between counterparty members; capital adequacy and balance sheet disclosure would be required only of net exposure in outstanding FX forward trades.

Source: Note on CCIL's website, Sept. 2007

Despite the GFC, the process of deepening FX markets continued. In 2008/09 futures were allowed and traded on exchanges.

Changes proposed in the draft guidelines announced in Paragraph 119 of the Second Quarter Review of Monetary Policy for the year 2009–2010, Reserve Bank of India included:

1. Importers and exporters with foreign currency exposures in trade transactions, permitted to write covered call and put options both in foreign currency–rupee and cross currency and also to receive premia.
2. AD Category-I banks permitted to offer plain vanilla cross-currency options to persons residents in India (other than AD Category-I banks), who transform their rupee liability into a foreign currency liability.
3. Given the facilities given in item 1 the facility of zero cost structures/cost reduction structures was to be withdrawn, since these opaque structures were used for speculation on rupee strengthening and imposed large losses on firms in 2008.

17 January, 2012: Hedging commodity risk on international exchanges allowed for listed companies through banks.

Reducing detailed oversight reduces transaction costs for firms, but the regulator has to prevent systemic risk. Although the strategy was to move from micro controls to regulating broad patterns, there was some back-tracking in times of high volatility. FX markets had deepened and the variety of hedging instruments increased, but the concern to increase the share of hedging transactions remained, as the assessment below from a market participant demonstrates:

Daily volumes in the currency futures market crossed US $4 billion in just over a year after the launch. But over 70 % of the volume traded came from jobbers and day traders. Open positions are an indicator of hedging. Banks and other players that arbitrage the OTC market accounted for another 12–15 %. So, the open interest on the market from hedgers and medium-term position-takers was only about 12–15 %. Compare this to the CME, where open interest averaged nearly 95 % of a daily volume of about $100 billion a day. For an emerging market currency like the Mexican peso, greater hedging volume takes the ratio of open interest to volume to 300 % (Mecklai 2010a). Since OTC transactions dominate, positions in futures markets alone are an insufficient gauge of hedging. But Rathinam and Aurora (2011) in a study of the off balance sheet activities of 15 Indian Scheduled Commercial Banks found that over 97 % of notional amounts assigned to derivatives were for trading not for hedging.

Over 2011–2013, there was some reversal in permissions due to global risk-off and excess rupee volatility and FX markets shrank somewhat.

2.2 Types of Intervention in FX Markets

Although the stated position remained the RBI would act to prevent excess volatility, markets were allowed to determine INR level and volatility subject to what remained of capital controls that were being reduced under domestic and international pressure. Intervention was temporarily suspended in 2007 at a time of strong inflows that made sterilization difficult, but resumed to accumulate inflows from October as the market stabilization bonds were negotiated for cost sharing with the government. The INR had to depreciate during post-Lehman equity outflows in order for them to take a write-down in asset values and share risk. The RBI did sell some reserves. Inflows resumed quickly, however, and up to end 2011, were just adequate to finance the CAD. So there was hardly any intervention in this period. This led to the market misperception that the RBI was unable to intervene in FX markets, aided by statements from the RBI about the large size of India's FX liabilities and potential capital movements relative to reserves. RBI communication was large-scale intervention was not possible since reserves even at 300 billion now just covered India's international liabilities. Just as policy allowed more market determination of the exchange rate becoming increasingly hands off, strong global risk-on risk-off in the period after the GFC created perverse movements in the exchange rate.

As inflows slowed due to global risk aversion after the Euro-debt crisis, markets shorted the rupee and it began to fall steeply, almost reaching 55. An environment of low growth and a rising CAD added to the fragility of FX markets. Measures to further liberalize inflows proved inadequate. So there was some reversal of liberalization—restrictions were put on FX markets and intervention resumed. RBI began to sell reserves in November 2011 as the INR spiraled downwards, and imposed restrictions on markets.

Retrospective taxation in budget 2012, and the Fed's taper announcement in May 2013, all led to outflows requiring RBI action[7]. Policy actions used included administrative measures such as controls, market restrictions, intervention or buying and selling in FX markets, signaling, and monetary policy measures such as the classic interest rate defense. Thus, it turned out that there were many feasible actions.

Table 3, which lists the policy measures taken over 2010–2014, attempts to assess their effectiveness by estimating this qualitative and quantitative impact on the exchange rate, that is, did a measure reverse or add to existing market movements and if so, by how much? The Table gives the basis points change in the INR/USD rate in the week before and the week after a measure. A negative entry implies an appreciation of the INR and a positive entry the reverse.

The Table indicates the most effective measure was the FX swap window[8] announced for oil marketing companies on 28th August, 2013. Not only did the INR strengthen substantially, but it reversed an existing depreciation. The peak value of INR/USD 68 was not regained. The rupee continued to appreciate after that, as other measures were added to the swap window that remained open till end-November. Measures that made more FX available, such as the subsidy for banks foreign borrowing or easier ECB also appreciated the INR. Restrictions on markets such as reducing position limits worked only sometimes, and total bans were not effective (see also Sect. 2.3).

Raising interest rates to defend the rupee in July 2013 was a total failure. The value of rupee depreciated from around 60 in July to a low of 68 in August. The 3 % rise in short-term rates was aimed at retaining debt flows since zero open positions[9] already prevented domestic banks from speculating against the rupee. Ten percent of the US $6.6 that had come in since 2011 left in June after the May taper-on announcement. Higher short rates did not stop the outflows and by November 40 % had left. Already high interest rate spreads and long-term rates rose, hurting the domestic recovery and domestic financial markets, where turnover fell further. Equity inflows, however, continued positive and were a healthy US $14 billion over 2011–2013. Debt flows also revived by September 2013, after short-term rates were normalized. Of the approximately US $50 billion FII inflows over 2013 and 2014, debt inflows were just a little over half.

---

[7]After zero intervention from January, monthly net purchases in USD million were 10678 over 2007:10 to 2008:10. This switched to net sales of 1505 over 2008:11 to 2009:4 as outflows intensified under the GFC. Average intervention was near zero at monthly net purchases of 285 over 2009:05 to 2011:10. But 2011:10 to 2013:07 saw heavy monthly net sales of 8580.

[8]By entering into fixed tenor sell/buy USD-INR swaps through designated banks, the RBI effectively lent dollars against rupees with the transaction to be reversed in the future as the companies returned the dollars.

[9]The net open position measures risks due to a banks' mix of buy and sell positions in different currencies. It is measured by the higher of net buy or net sell positions across all currencies. A zero open position means a bank cannot have foreign currency assets exceed foreign currency liabilities in its balance sheet or have an unsettled buy position in foreign currency. This reduces selling pressure on the rupee coming from banks.

Signals that the RBI was unable to intervene and the INR should be left to the markets in 2011 had a large counter-productive impact. Well-designed signals had the desired effect, as with the new RBI governor's joining and speech on September 4, 2013, announcing subsidies for banks raising FX deposits. Fed announcements also impacted the INR. It appreciated after the US Fed's 18th September postponement of the taper.

The lessons from this experience were the importance of designing policy in line with the current state of capital account convertibility, restraints on debt flows, and evolution of markets. Given India's growth prospects and relatively greater reliance on growth driven-equity flows, the interest rate defense was counter-productive and could have been avoided. The value of equity investors' assets decreases with a sharp depreciation, but an ineffective interest rate defense does not help existing equity investors, even as reduced growth harms new entrants. Even debt flows respond to risk premiums determined by overall macroeconomic stability, not just to narrow interest differentials.

Under adverse expectation-driven outflows, the market demand and supply for FX will not determine an exchange rate based on fundamentals. Smoothing lumpy foreign currency demand in a thin and fragile FX market is important. Direct provision of FX to oil marketing companies was first used in the mid-1990s.[10] It is a useful way to provide FX reserves to a fragile market without supporting departing capital flows. It also encourages domestic entities to hedge. It showed there are innovative ways of using reserves, which can be built up again during periods of excessive inflows. Although swaps add exchange rate risk to the RBI's balance sheet, it need not materialize over the short life of the swap if markets are successfully calmed.

In general, intervention must not be one-sided and has to be strategic, drawing on CBs superior aggregate market information. Timing is very important and must be based on market intelligence covering net open positions, order flow, bid-ask spreads (when one-sided positions dominate dealers withdraw from supplying liquidity and spreads rise), turnover, and share of interbank trades. Emerging markets (EMs) typically have less information and more uncertainty, so signaling can be effective. A variety of signals can be used.

It is only if these polices are not effective that restricting markets may become necessary. But that should be avoided, to the extent possible, since it has adverse side-effects. Modeling strategic interaction between differentially informed speculators and the Central Bank, with EM features included shows why types of intervention and signaling that are not normally effective, may work in FX markets like those in India.[11] Section 3 shows why bounds on exchange rate volatility can reduce noise trader entry poorly informed traders and thus improve policy autonomy. An EM is likely to have a larger share of such traders.

---

[10] I thank Dr. Y.V. Reddy for this point.

[11] This analysis is based on Goyal et al. (2009).

Each party is assumed to make inferences based on the other's behavior, under shocks that affect information extraction. The speculative demand curve is downward sloping in the spot rate and stable if there is greater uncertainty about fundamentals, or if speculators prior on the target and CB's weight on the target are small. The first normally holds in an EM, and the second and third hold if the target is diffuse. The estimated speculative demand using data over 2002–2008 was found to be downward sloping. CB purchase of dollars tended to depreciate the domestic currency, and reduced its volatility. Expectations were stabilizing and not perverse.

The results suggest a CB in an EM can optimally reveal some information, but should not announce an explicit target, or reveal its trading tactics. There is a range of possible signals. Greater uncertainty about fundamentals makes it more worthwhile for the CB to reveal some information about an exchange rate target.

As the model implies, market microstructure variables were found to affect intervention efficacy. Market microstructure variables such as merchant and dealer net demand (order flow) and turnover variables affected the level and volatility of exchange rates. Merchant turnover was a driving force, volatility and expected volatility increased dealer turnover. It follows the microstructure variable matter and policy has a role in limiting volatility.

Anticipated intervention decreased turnover, so expectations from intervention were stabilizing and not perverse. They dampened rupee volatility. Since markets form expectations of intervention activity and respond strategically to it, more transparency may reduce the scope for such arbitrage. Although more transparency could reduce speculative positions, a diffuse target may work better. Estimated strategic market behavior and model derivations both indicated intervention and signaling to be an effective influence on exchange rates in the Indian context.

The interest rate differential had weak effects on the exchange rate but strong effects on market turnover. With the extant level of controls, the effect of the interest rate on the domestic cycle was stronger than its effect on the exchange rate. Given the evidence of the impact of policy on markets, more transparent intervention may effectively influence exchange rates in the Indian context, leaving interest rates free to target the domestic cycle. Next, we turn to examine the efficacy of market restrictions in more detail.

2.3 Impact of Measures on Domestic Markets

The repeated scams and financial mishaps of the 1990s demonstrated the fragility of a controlled system. Therefore, financial reforms towards steady market deepening were undertaken. But the global financial crisis demonstrated the wisdom of India's slow and steady approach to market liberalization, and the necessity of prudential regulation to reduce risk-taking. Action on the INR was not, however, always consistent with these lessons. Sometimes actions were too hasty and cautious steps forward to deepen domestic markets were reversed, but did not always succeed in reducing rupee volatility.

On December 15, 2011 the RBI reduced banks net open position limits (NOPL) by 75 %. This forced MNCs banks to close their huge long USD positions. Bank boards did not want to have to report a fine paid to RBI for noncompliance. On representation the RBI clarified that genuine trade-based positions would be allowed. They also canceled rebooking of forwards (corporates were shifting hedging band to 55–60, paying the option fee of 2 %). These measures to reduce speculation by exporters and banks appreciated INR/USD from 54.2 to 52.7 (Table 3). Rupee was back at 50 within a month.

Adverse tax measures in the March 2012 budget triggered outflows again and the INR again reached 55. Netting of positions in currency futures/options with over-the-counter (OTC) positions was disallowed and position limits of banks further reduced in May 2012 but the next week saw the INR further depreciate by 0.90. Over June and July 2013 foreign banks open positions in INR/USD was reduced to almost zero and any proprietary activity by banks in currency futures was banned, but depreciation continued.

Figure 1 shows various market restrictive measures reduced market turnover sharply in the currency derivatives markets in exchanges, while total turnover including the dominant over-the-counter (OTC) FX trading in banks also fell. This suggests the two types of markets are complements rather than substitutes. Banks are some of the largest traders on exchanges. Exchanges are thought to be dominated by speculative position-taking since no real underlying is required unlike in the RBI regulated OTC markets. But in FX markets, worldwide portfolio rebalancing types of transactions between market makers are normally much larger than those based on real exposures. These allow banks, as well as small firms that may not get a good deal at banks, to lay-off risks in futures markets. Future markets are more transparent and equitable but expectations are especially important in such markets and can lead to one-way positions, so prudential regulations are important.

The other effect of restrictions was offshore markets grew at the expense of domestic markets. In more open regimes, restricting domestic markets encourages transactions to migrate abroad. Though an accurate assessment of the volumes is difficult, estimated daily INR NDF turnover was around US $100 million in 2003/2004 and grew substantially since. By April 2013, it exceeded onshore

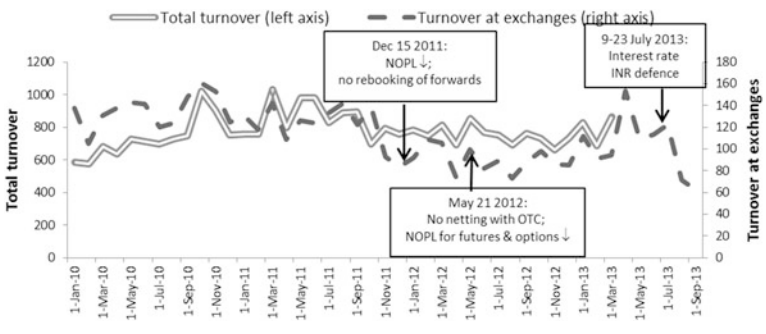


Fig. 1 FX market turnover (USD billion)

Table 4 Rise in cross border derivatives

Total outright forwards						
	Domestic: net-gross basis			Global INR: net-net basis		
	$m	Cross border %	% of spot	$m	Cross border %	% of spot
Apr-13	3,743	14.64	24.19	24,395	60.43	160.23
Apr-10	4,895	8.56	36.50	13,620	52.11	100.69

Note Net-gross basis adjusts for local inter-dealer double counting. Net–net basis adjusts for local and cross-border inter-dealer double counting. BIS (2013) warn turnover for years prior to 2013 may be under-reported, especially for EMs

trading.[12] Onshore market affects price discovery in INR NDF market. However, in volatile market conditions NDF markets influence spot and forward onshore markets. A large spread between INR NDF rate and INR futures/forward rate impacts the spot rates significantly. The INR forward rate is influenced by the movement of INR NDF futures and spot rates with some lag (Behera 2011). Although such markets create problems for policy, they normally wither away as domestic markets deepen (Ma et al. 2004).

Although they are difficult to measure precisely BIS (2010 Table E6, 2013 Table 6.3) shows OTC INR turnover (net-gross basis) outside the country rose from 50 % (20.8 USD billion) in 2010 to 59 % (36.3 USD billion) in 2013 of the total turnover. Table 4 shows the sharp rise in INR forwards, which include NDF, between 2010 and 2013 when domestic restrictions were imposed. The rise as well as much higher net-net compared to net-gross value shows much of the growth was abroad.[13] A rising share of the non-deliverable forward market is against the objective of developing and deepening domestic markets. Moreover, domestic regulators are unable to influence offshore markets. Therefore, using prudential regulations instead of forbidding transactions, would also have the advantage of not driving markets overseas.

2.4 Encouraging Hedging

A conviction of possible two-way movement of the exchange rate, large enough to deliver a substantial loss to one-way bets is a prerequisite for hedging or the laying off of currency exposure. Despite deepening FX markets, the moderate two-way movement within an implicit 5 % band seen over 2004–06 was not sufficient to

[12]Mecklai (2011) argues it had become higher even by 2011.

[13]It is noteworthy that the relative size of forwards in net-net global GBP trade and net-gross trade in UK is reversed for the deepest FX market, UK. In April 2013, the net-gross at US $309 billion was much larger than net-net at US $69 billion. FX trade in the UK is very large in currencies other than the GBP, including the INR. It follows a large share of transactions involving the INR occur abroad. Relative turnover sizes for other EMs are like that for India.

overcome strong expectations of medium-term appreciation given India's high growth rate. In 2007, market expectations of the INR/USD rate had even reached 32. Many corporates borrowed abroad based on such expectations, increasing currency risk. Some had entered into so called hedging deals, which were actually bets on the value of the Swiss Franc. With the volatility in currency markets and steep rupee depreciation in 2008 many firms lost money. Many such deals, where Indian banks were often a front for foreign banks, sidestepped existing rules that prevented leverage or underlying risk that exceeded export income. Although firms were not allowed to write options deals were structured so that in effect firms were writing options. The deals were so complex that firms sometimes did not understand what risks they were taking. After post GFC episodes of excess volatility, the rupee was managed and stayed in a tight band of INR/USD 64–66 over September 2013–December 2014. Since international interest rates were much lower than Indian rates firms were again tempted into unhedged foreign borrowing, although such borrowing, for example, through ECB, was capped. Even so, gross ECBs worth around US $264.4 billion came in from 2001 until Oct 2014. Refinancing takes up a large share.[14] Not hedging is dangerous since the bulk of the borrowing was by infrastructure firms that do not have any natural hedge in the shape of exports.

Thus, availability of more instruments, alone, only makes leveraged speculation, or bets on future currency value, possible. Establishing inducement to hedge through sufficient flexibility of the exchange rate, along with transparency, clarity and information, and strategic use of controls, are more important (Shiller 1993; Shefrin 2002). Completing markets will not by itself reduce speculation. Incentives have to be changed, and better information provided on fundamentals.

3 FX Markets and Policy

Well-designed signals can help use the structure of FX markets with varying trader information levels to achieve policy objectives such as reducing exchange rate volatility. An EM is likely to have a larger share of poorly informed traders.

Jeanne and Rose (2002) have a model with informed and noise traders (n) in an FX market. The benefit of entry for noise traders rises with excess returns or risk, ρ, but falls with the variance of the spot rate, var(S). But both ρ and var(S) are functions of n. A rise in n lowers ρ since a larger number of traders are demanding the currency thus spreading risk; but raises var(S) since entry raises volatility and therefore risk. Both reduce entry. But the rise in var(S) itself raises ρ, thus increasing entry. Noise traders therefore have two counteracting roles. They both create risk and share risk, making multiple equilibria possible.

[14]For example, of the US $2.8 billion that came in October 2014 28 % was for refinancing.

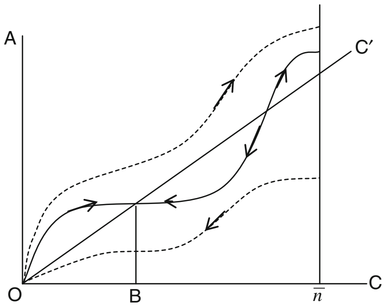


Fig. 2 Multiple equilibria in FX markets

This can be shown in a simple diagrammatic device following Goyal (2006). $G(n)$, the function giving the returns to noise trader entry is graphed against n in Fig. 2. The two opposing effects of n on $G(n)$ give the curve the shape shown in the Figure (derived in the appendix). At low n G is high since var(S) is low. It falls as more n share risk reducing ρ. But further entry creates risk raising var(S) and ρ. At high n G rises again, as the rise in ρ dominates that in var(S).

The lowest dashed curve is the case where fundamentals are strong. Therefore, excess returns are so low that it is not worthwhile for any noise trader to enter. Point O is a stable equilibrium, as is C', on the upper dashed curve. On this curve, fundamentals are so weak that all the noise traders enter. Multiple equilibria occur for intermediate fundamentals when $G(n)$ is such that it cuts the 45° line. These are equilibria since $G = n$ at these points. A is stable since for a small departure from A, net returns are such as to bring entry back to n^*. C is also stable. But B is unstable. Here the curve cuts the 45° line from below.[15] So, a small rise in n raises returns and induces more entry until C is reached. A small fall in n reduces returns and entry until A is reached.

At C' and C all noise traders enter. At A, var(S), ρ and therefore entry n are all low; at C var(S) and ρ are both high and maximum entry $\bar{n}$ takes place. Between AB, since $G(n) < n$, entry falls back to n^* at A; between CB, since $G(n) > n$, entry rises and n increases to $\bar{n}$ at C. Therefore, O, A, C and C' are stable equilibria, while B is unstable. It follows exchange rate volatility can be low if fundamental variance is low, but it can be either high or low for intermediate levels of fundamentals.

Monetary policy can improve welfare if it leads to the selection of low volatility equilibria. It can do this by committing to var(S) $\leq v$. Then fewer n enter, this decreases var(S) leading to the selection of the low entry equilibrium A. This is a policy-free lunch since markets help monetary policy achieve the reduced volatility it had committed to.

[15]The style of proof is similar to the well-known Keynesian cross where the aggregate demand line cuts the 45° aggregate supply line from above in a stable equilibrium. The intuition is similar in the fixed point theorems used in general equilibrium theory. The proof following Goyal (2006), a major simplification of that used in Jeanne and Rose (2002), is given in the Appendix.

A stable exchange rate regime, where policy constrains var(S) to be less than or equal to v, leads to lower entry of noise traders taking the economy to an equilibrium with low exchange rate variance.[16] Although there is a loss of monetary autonomy in adopting the restriction on var(S), the loss is of second order, as the level of noise in the economy is decreased. Monetary policy response function is constrained out of equilibrium, but there is no sacrifice of monetary autonomy in equilibrium. In the post GFC period, Indian exchange rate volatility was allowed to rise too much, while there were also periods when it was too low. Goyal and Arora (2012) give evidence CB speeches affect the exchange rate but over the period signals were poorly used. Table 1 shows this in 2011 but indicates improvement in 2013.

4 Conclusion

The chapter gives an overview of international and Indian FX markets and discusses market microstructure concepts such as 'order flow', bid-ask spreads, limit order, novation, netting, settlement and counterparty risk, electronic broker and crossing systems, inventory risk and 'hot potato' trading. It also records the very large size of FX transactions and the growth and deepening of Indian FX markets aided by steady regulatory changes. OTC and bilateral transactions dominate in FX markets, but it is not yet clear if technological changes and the international post global crisis regulatory emphasis on transparency will lead to more exchange trading. It may be that FX markets are fundamentally different, with the bilateral market structure essential for stability.

It also draws out the implications of structure for CB intervention, signaling, and exchange rate policy especially in an EM. Committing to a low exchange rate volatility regime provides something like a free lunch to monetary policy, because, as the analysis in the last section showed, entry of noise traders is reduced leading to lower volatility. In an EM, there tends to be greater uncertainty about fundamentals. This tends to make the speculative demand curve well behaved under strategic interaction between differentially informed speculators and the Central Bank (CB). A diffuse exchange rate target and strategic revelation of selected information can be expected, therefore, to be effective. Indian experience supports these results.

Post GFC experience even delivers an effectiveness ranking of policy actions influencing the exchange rate at the current state of capital account convertibility. The most effective is to address fundamental weaknesses that can trigger adverse expectations, but second, reserves and signaling can be used to smooth market demand and supply keeping the exchange rate within bounds. The efficacy of

[16]Such an induced entry of noise traders was illustrated by the large-scale shorting of the INR in December 2011 after the CB's communications were taken to imply it would not intervene to support the rupee.

building-up reserves, using them in targeted intervention, and of credible communication to markets was clearly demonstrated. Since capital flows do not always match the net import gap, a CB should be ready to close any short-term demand supply mismatch. Reserves can be built up again during periods of excessive inflows.

Prudential measures such as reducing open positions worked better than a ban on a market or a transaction-type. But it reverses attempts to deepen markets and leads to a migration of activity abroad. If market restrictions become necessary, therefore, they should be carefully targeted. Even if credit curbs are used they should apply only to specific commodities such as gold imports. Since these administrative measures reduce one-way positions, a general liquidity squeeze such as an interest rate defense that hits other markets, should be avoided. EMs should continue the push for greater global policy coordination and measures that reduce the financial over-leverage that leads to capital flow volatility, even while developing regional safety nets.

Acknowledgment I thank Akash Kumar Baikar for research and Reshma Aguiar for secretarial assistance.

Appendix: Deriving Equilibrium Noise Trader Entry

Equilibrium requires that a constant number of noise traders, n, enter. Noise trader's benefit from entry rises with ρ and fall with var(S). Entry will occur only as long as this benefit exceeds their cost of entry. Equation (1) defines an implicit, smooth twice-differentiable benefit function:

$$B(\rho, \mathrm{var}(S)); \quad B'[\rho] > 0, \quad B'(\mathrm{var}(S)) < 0 \tag{1}$$

Where a superscript dash indicates a partial derivative. Trader j will enter the market as long as:

$$B(\rho, \mathrm{var}(S)) \geq c_j \tag{2}$$

But both ρ and var(S) are functions of n. Equilibrium ρ, equates demand to supply in the domestic currency security market. It is given by Eq. (3), written implicitly as:

$$\rho^* = \rho(\mathrm{var}(S), n); \quad \rho'(\mathrm{var}(S)) > 0, \quad \rho'(n) < 0 \tag{3}$$

A superscript * denotes an equilibrium value. Similarly, the equation for equilibrium var(S) is written implicitly as:

$$\mathrm{var}(S)^{*} = \mathrm{var}(S)(n); \quad \mathrm{var}(S)'(n) > 0 \tag{4}$$

In equilibrium either all noise traders will enter, or none will enter, or some will enter, so that $n \in [0, \overline{n}]$. If $B() > c_j$ for all noise traders, all will enter. If $B() < c_j$, no noise trader will enter. In an equilibrium with interior values, (2) will hold with equality, and $\overline{\rho}^{*}$ and $\mathrm{var}(S)^{*}$ will take critical values such that the marginal noise trader is just indifferent to entering.

$$B(\rho^{*}, \mathrm{var}(S)^{*}) = c_j \tag{5}$$

At $\rho < \rho^{*}$ or $\mathrm{var}(S) > \mathrm{var}(S)^{*}$, benefits to entry are lower than at equilibrium so n will shrink. Since both ρ and var(S) depend on n, a function $G(n)$ can be defined, that determines entry: $G(\rho(\mathrm{var}(S), n), \mathrm{var}(S)(n))$. If $n \neq G(n)$ it cannot be an equilibrium. Hence equilibrium entry is:

$$n^{*} = G(\rho(\mathrm{var}(S), n^{*}), \mathrm{var}(S)(n^{*})) \tag{6}$$

If $B() > c_j$ then $n < n^{*}$, noise trader entry will occur and n will rise. Since ρ falls with n but rises with var(S), and var(S) rises with n, multiple equilibria are possible. $G'(\rho) > 0$ and $G'(\mathrm{var}(S)) < 0$, therefore although $G(n)$ can be high since var(S) is low it falls with n at low n as ρ also falls and decreases $G(n)$. The risk sharing function dominates. But at high n, the positive effect of n on var(S) and therefore on ρ will dominate—ρ will rise as risk rises. Hence G(n) will also rise with n at high n. Therefore equilibria are possible both at low and at high n. Either a few or a large number of noise traders will enter the FX market. But, in each equilibrium n takes a fixed value, given by the function G(n). Noise traders create risk so var(S) rises and ρ falls with their entry (n). But ρ also rises with var(S), since they also share the risk they themselves create.

References

Behera H (2011) Onshore and offshore market for Indian rupee: recent evidence on volatility and shock spillover. Macroecon Financ Emerg Mark Econ 4(1):43–55

BIS (Bank of International Settlements) (2013) ‘Triennial Central Bank Survey of foreign exchange and derivatives market activity in 2013’. December, and earlier reports BIS (2007, 2010). http://www.bis.org/publ/rpfx13.htm

Geithner TJ (2004) Hedge funds and their implications for the financial system. Federal Reserve Bank of New York. http://www.newyorkfed.org/newsevents/speeches/2004/gei041117.html

Goyal A (2006) ‘Exchange rate regimes: middling through’. Glob Econ Rev 35(2):153–176

Goyal A (2015) ‘External shocks’, Chapter 3. In Mahendra Dev S (ed) India development report 2015, IGIDR and Oxford University Press, New Delhi, pp 36–51. (Earlier version) http://www.igidr.ac.in/pdf/publication/WP-2014-046.pdf

Goyal A, Arora S (2012) ‘The Indian exchange rate and central bank action: an EGARCH analysis’. J Asian Econ 23(1)

Goyal A, Nair RA, Samantaraya A (2009) 'Monetary policy, forex markets, and feedback under uncertainty in an opening economy'. Development Research Group, Department of Economic Analysis and Policy, Mumbai, Study No. 32. http://rbidocs.rbi.org.in/rdocs/Publications/PDFs/DRGMP030909.pdf

IMF (2008) 'India: selected issues'. IMF Country Report No. 08/52 January. International Monetary Fund, Washington

Jeanne O, Rose AK (2002) Noise trading and exchange rate regimes. Q J Econ 117(469):537–570

Lyons RK (2001) The microstructure approach to exchange rates. MIT Press, Cambridge, MA

Ma G, Ho C, McCauley RN (2004) 'The markets for non-deliverable forwards in Asian currencies'. BIS Q R 81–94.

Mecklai J (2010a) 'Market maniac: bringing hedgers to the futures market, Business Standard. http://www.business-standard.com/article/opinion/jamal-mecklai-bringing-hedgers-to-the-futures-market-110010800002_1.html. Accessed 2 Jan

Mecklai J (2010b) 'Market maniac: The OTC-FX market—Falling behind the curve', Business Standard. http://www.business-standard.com/article/opinion/jamal-mecklai-otc-fx-market-falling-behind-the-curve-110100100093_1.html. Accessed 1 Oct

Mecklai J (2011) 'Market maniac: Has RBI lost control of the rupee?', Business Standard. http://www.business-standard.com/article/opinion/jamal-mecklai-has-rbi-lost-control-of-the-rupee-111070100044_1.html. Accessed July 1

Rathinam FX, Arora D (2011) 'OTC derivatives markets in India: recent regulatory initiatives and open issues for market stability and development'. Macroecon Financ Emerg Mark Econ 4 (2):235–261. http://www.tandfonline.com/doi/pdf/10.1080/17520843.2011.580571

RBI (1995) 'Report of the expert group on foreign exchange markets in India' (Sodhani Committee Report). Reserve Bank of India, Mumbai

Sager MJ, Taylor MP (2006) Under the microscope: the structure of the foreign exchange market. Int J Financ Econ. 11:81–95

Sarno L, Taylor MP (2002) The Economics of Exchange Rates. Cambridge University Press, UK

Shefrin H (2002) Beyond greed and fear: understanding behavioral finance and the psychology of investing. Oxford University Press, Oxford

Shiller RJ (1993) Creating institutions for managing society's largest economic risks. Clarendon Lectures in Economics, Clarendon Press, Oxford

Tarapore SS (2006) Report of the committee on Fuller capital account convertibility. Reserve Bank of India, Mumbai

Global Foreign Exchange Market: A Crisis Analysis

Gagari Chakrabarti

What we know about the global financial crisis is that we don't know very much…

Paul Sammuelson

Abstract The global financial market is often taken as an inherently fragile system that is prone to irrational exuberance, unfound pessimism and crises. Faced with such a system, researchers and analysts often seek to explore the roots of crises and the channels through which they reverberate from the center to periphery. While financial disasters are probable it would be nice if they are predictable. The investors, analysts and the policy-makers would sit comfortably if stress or crisis in one financial market could be predicted from those in other markets. This is particularly the area where the present study intervenes with a focus on the foreign exchange market. It considers three exchange rates defined between (i) two emerging nations (India and Singapore), (ii) an emerging and a developed nation (India and the US) and (iii) two developed nations (the US and the UK). In terms of stress indexes defined for each of these markets, it found no causality between stock market and foreign exchange market stresses for the developed-developed market pair. For the emerging markets, particularly for India, such channels of stress transmission remain and foreign exchange market crisis and stock market crisis (whether generated domestically or emanating from the developed, foreign market) may appear as "twin". For Singapore, however, such a channel exists where stress is generated only in the other emerging market. Thus, the emerging markets that experience huge inflow of foreign capital in their stock markets might take stock market crises and foreign exchange market crises as twin. The policy implications, however, might differ. In some cases, it would be enough to regulate the domestic stock market, but in some other instances crises may be contagious coming from stock markets abroad.

The author acknowledges discussion with Dr. Chitrakalpa Sen, Assistant Professor–Economics, BML Munjal University, Haryana, India.

G. Chakrabarti (✉)
Presidency University, Kolkata, India
e-mail: gagari.econ@presiuniv.ac.in

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_24

Keywords Global financial market · Financial crisis · Stress index · Granger causality · Conditional correlation

1 Introduction

Whilst each cloud comes with a silver-lining, every silver-lining has a cloud behind it. Each boom brings a bubble in its train; and every crisis starts with a bubble. Fed by the initial and often irrational euphoria, financial booms soon tend to develop into bubbles: a trouble-spot that usually goes on unnoticed or is left unattended in intense speculative interest. Even in the presence of irrational exuberance, sheer recklessness or fraudulent behaviour people seem to keep absolute faith on the four most dangerous words in finance—'*this time it's different'*. Even in perceptible absence of real economic value creation, intervention is claimed to be sub-optimal. At one point in time financial vulnerabilities surpass any tolerance level and bubble transforms itself from a state of mind to a driving force for economic change. The bubble bursts, panic spreads with an ultimate devastating impact on other sectors of the economy. Frequent crashes in the financial market have raised questions about the myth of the *efficient market.* Shiller has pointed out that "*while markets are not totally crazy, they contain quite substantial noise, so substantial that it dominates the movements in the aggregate markets*" (Shiller 2003). The global financial market should be taken as a system prone to "irrational exuberance and unfound pessimism. It is, in other words, extremely fragile and prone to collapse" (Roubini and Mihm 2010).

In such an inherently unstable global financial system, researchers and analysts are often interested in exploring the root of such crisis and the channels through which crisis reverberate from the center to periphery. While financial disasters are probable it would be nice if they are predictable. The investors, analysts and the policy-makers would be equally happy if stress or crisis in one financial market could be predicted from those in other markets. This is particularly the area where the present study intervenes with a focus on the foreign exchange market crisis. A crisis in the foreign exchange market is often thought to transpire from a wide and complex variety of economic factors. In this era of financial integration, these are often triggered, particularly in an intrinsically vulnerable foreign exchange market, by crises in other financial markets nearby. The newer generation models of twin crises emphasize the role financial sector and capital flows in currency crises (Stoker 1995; Mishkin 1996; Velasco 1987; Calvo 1995; Reinhart and Vegh 1996; Kaminsky and Reinhart 1999). The present study delves in the issue to explore whether crises in the foreign exchange market could be predicted from the crises in other financial markets, particularly the stock market.

2 Earlier Literature in the Field

Owing to the great amount of integration between financial markets, volatility in foreign exchange is often considered to directly affect the stock markets. Contagion or co-movement between the financial markets has long been a hot topic for discussion. There have been numerous studies addressing this causal or predictive relationship between stock and foreign exchange market and most of them found a direct correlation between these two markets (Granger et al. 2000; Mishra 2004). Studies have shown that stock market responds to volatility in foreign exchange market or in other words, there is a significant amount of volatility spillover between stock and foreign exchange markets. A change in the exchange rate directly affects both firms who are involved in international trade through their foreign exchange exposure as well as firms who trade only domestically through their portfolio adjustments (Bahmanee-Oskooee and Sohrabian 1992). This in turn, changes the share prices of the affected firms and hence affecting the stock market. Nath and Samanta (2003) in their study in context of Indian Rupee/US Dollar, used cointegration test and Granger causality to conclude that there is a causal relationship from stock market returns to foreign exchange returns and only a mild causality in the reverse direction. Ajayi et al. (1998) took seven developed economies and eight emerging Asian markets and used Granger causality to test for dynamic linkages. While most the developed economies showed a causal relation from foreign exchange to stock market, the result for the Asian economies were largely mixed. Muhammad and Rasheed (2002) studied the relationship between stock and foreign exchange markets in context of Pakistan, India, Bangladesh and Sri Lanka using cointegration, Vector Error Correction model and Granger causality. They found no short-term association between these countries' foreign exchange market and stock markets. However, Bangladesh and Sri Lanka exhibit a long-term relationship between these two markets. Stavárek (2005) conducted his study in context of nine countries (Austria, France, Germany, Slovakia, Czech Republic, France, Hungary, the UK and the US) and found significant short-term and long-term causal relations from stock market to foreign exchange markets for all of them. Tabak (2006) conducted an extensive study in context of Brazil and found strong linear Granger causality from the stock market to foreign exchange market and nonlinear Granger causality from the foreign exchange market to stock market. Pan, Fok and Liu (2007) found evidence of causal relationships between daily exchange rates and stock prices for Hong Kong, Japan, Malaysia and Thailand before the 1997 Asian financial crisis. A causal relation from the stock market to the foreign exchange market was also found for Hong Kong, Korea and Singapore. No country exhibited causality from stock prices to exchange rates during the Asian crisis, while exchange rates were found to influence stock prices for all countries except Malaysia. Several other researchers have shown similar dynamic linkages from stock market to foreign exchange market. Kasman (2005) in context of Turkey, Horobet and Ilie (2007) for Romania, Yang and Doong (2004) for G7 countries, Shew (2008) for Singapore, especially after the Asian crisis, Wickremasinghe (2006) for Sri Lanka, Choi et al. (2010) for New Zealand are some of the significant studies.

The present study explores whether the crisis in the foreign exchange market could be predicted from the movements in other financial markets, particularly the stock market. Specifically, it inquires whether and to what extent the possible stressed situations in foreign exchange market could be predicted from the stresses in the stock market. For that purpose it seeks to define a stress index for each market and conducts the analysis in terms of these stress indexes. In this way, the study is different from the earlier approaches in the field that emphasize simply on the interdependence between stock prices and foreign currency rates.

3 The Study

While dealing with the issue, the study takes up the following trajectory and tries to answer the following set of questions:

1. Have the movements in the foreign exchange rates and stock prices been similar over the period of time?
2. Mere similarity in movements does not indicate any association. Hence, to what extent these markets move in conjunction?
3. However, mere correlation between returns may not imply correlated stress across markets. Prediction of stress in one market from that in the other is even more difficult (even in correlated markets) simply because of the fact that all change in market prices do not inflict stress on it. Hence, specifically, how to define stress?
4. Finally, does the stress in one market cause that in others? And, if so, is the relationship independent of the levels of financial development across markets?

Over a period of 15 years ranging from 2000 to 2014, the study selects three exchange rates defined between (i) two developed financial markets namely the US and the UK; (ii) one developed (the US) and one emerging (India) market; and (iii) two emerging markets, namely India and Singapore. Thus, the three exchange rates chosen are British Pound per unit of US Dollar, Indian Rupee per unit of US Dollar and the Singapore Dollar per unit of Indian Rupee. The study would henceforth describe these rates as GBP/USD, INR/USD and SGD/INR respectively. These currencies are primarily floating. However, central government might intervene in extreme situations to avoid excessive and undesirable appreciations or depreciations.

The choice of the US and the UK in the first category is reasonably justified. The highly developed financial systems in the UK and the US have made these two the largest individual markets in the global arena. The widely and hugely traded GBP/USD ranks third among the most traded currency pairs in the global foreign exchange market. The trade in GBP/USD currency pair comprises nine percent of the total daily trading volume. In the second category, Indian currency and its relationship with a major global currency remains our point of concern. Although Indian currency is yet to be a full-float, the INR/USD trade is having a mass participation. Since late 2011, currency pairs between USD and currencies from BRIC nations, particularly the INR/USD are experiencing rapid growth. By late 2012, daily volume of trade in USD/INR currency pair stood at USD 2.4 Billion. In between October 2011 and October 2012, the volume of trade increased by 392 % (Source: BIS, AITE Group 'Global FX Market Update 2013: Increased Market Transparency, More Competition', June 2013). Singapore is one of the emerging markets in the Asian region. According to the Bank for International Settlements, the Singapore Dollar was ranked twelfth among the most actively traded currencies in the foreign exchange market in April 2010. It accounted for nearly 1.4 % of average daily foreign exchange market trading volume. Singapore dollar is actively traded against the Indian Rupee.

The study considers two stock markets for each of these exchange rates. For GBP/USD rates, it selects the Dow Jones Industrial Average and the FTSE 100; the two significant indexes from the US and the UK respectively. For INR/USD currency pair, it selects BSE SENSEX (India) and Dow Jones Industrial Average (US). For SGD/INR currency pair, it considers BSE SENSEX from India and FTSE Straits Time Index from Singapore. The most referred and closely watched Dow Jones Industrial Average, invented in 1896, is a price-weighted index of thirty most significant 'blue-chip' stocks traded on the New York Stock Exchange and the NASDAQ. The index includes stocks from sectors such as conglomerate, consumer finance, telecommunication, aerospace and defence, construction and mining equipment, oil and gas, software and computer networking, beverages, chemical industry, banking, financial services, pharmaceuticals, retail, entertainment and insurance. By November 2014, the float-adjusted market capitalization of the thirty-stock index stood at USD 4864.4 Billion. The one-year and ten-year annualized total return has been 13.42 and 8.27 % respectively (www.djindexes.com). The FTSE 100, introduced in 1984, is a market-capitalization weighted index of 100 blue-chip stocks traded on the London Stock Exchange. This index is seen as a barometer of business prosperity in the UK Market. The index includes stocks from sectors such as Banking and Insurance, mining and engineering, healthcare, consultancy, oil and gas, defence, property, media and entertainment, IT and telecom, industrial products, fashion and consumer goods, food, power and energy, chemicals, transportation and retail. By November 2014, the net market capitalization stood at GBP 1,704,341 million. It has yielded an annualized return of 10.8 and 9. 1 % over the last three and 5 years respectively (www.ftse.com). The S&P BSE SENSEX, introduced in 1986, is a free-float market-weighted stock market index of thirty blue-chip companies listed on Bombay Stock Exchange. It considers companies from different sectors such as capital goods, consumer durables, finance, FMCG, healthcare, real estate, IT and telecommunication, metal, metal products and mining, oil and gas, power, transport equipment and others. By November 2014, the market capitalisation of BSE SENSEX touched Rs. 100 Lakh Crore. It is taken as one of the world's top twenty stock exchanges by market capitalization.

FTSE Straits Times Index (STI), introduced in 1966 and revamped and re-launched in 2008, is the market capitalization-weighted benchmark index of the Singapore stock market. It is constituted of the top 30 companies listed on the Singapore Exchange. As of February 2014, the market capitalization of the index stood at 262,718 million Singapore Dollars. It includes companies from sectors such as banking and insurance, telecommunication, real estate and holding, airlines, beverage, consumer goods, engineering, marine, diversified industrials, transportation and tourism, oil, farming and fishing, and others. Over the past year, the index offered a return of 9.52 % (www.Bloomberg.com).

3.1 Stress in the GBP/USD Exchange Rate and the Two Stock Markets

The study starts its inquiry by plotting the exchange rate series and the two stock index series against time. Figure 1 shows the movement in GBP/USD exchange rate series over the 15-year-period ranging from January 2000 to December 2014. A major drop in the series is visible in January 2008. The steep fall continued for a period of one year when the series hit the slump in January 2009. The series is otherwise characterized by fluctuations that rarely developed into crisis.

Figure 2 shows the movements in the US and the UK stock markets over the same time period.

The movements in the two stock markets have been quite similar over the chosen time period. Both the markets turned bullish since 2003 and slipped in January 2008. Since January 2009, both the markets are rising steadily. The upturn in the US market has been more significant than the UK market in recent years. The movements in the foreign exchange market have been similar to these but such a graphical introspection is hardly sufficient to predict stress in foreign exchange market from those in the stock markets.

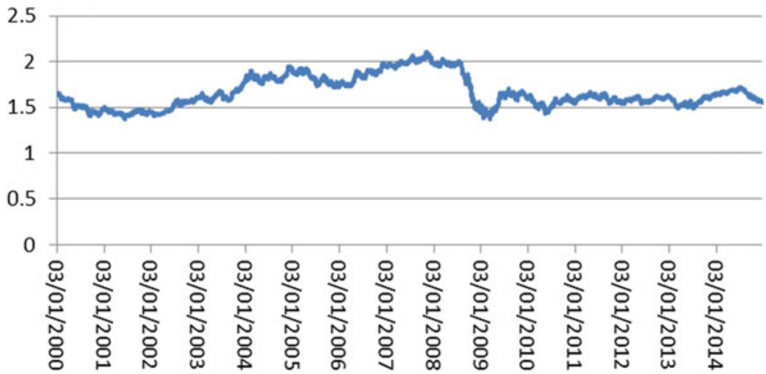


Fig. 1 Movements in GBP/USD series (2000–2014)

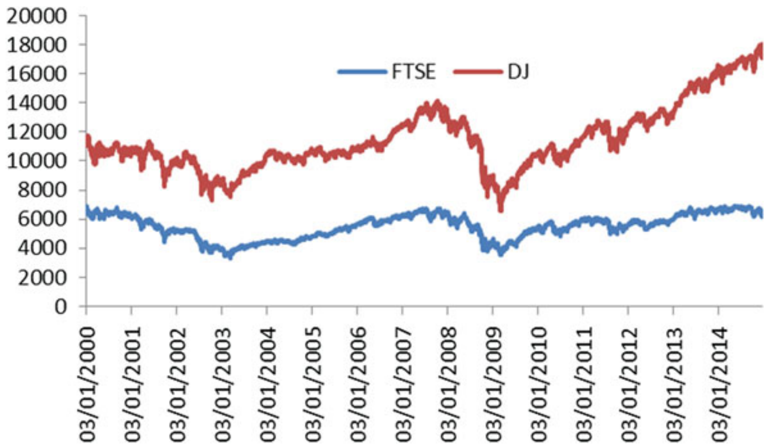


Fig. 2 Movements in FTSE 100 and DJIA series (2000–2014)

To explore predictability of foreign exchange market stress from stock market crises, we start from the exploration into possible interconnection between foreign exchange market and the two corresponding stock markets. The daily change in each of the series are calculated using the formula $R_t = \ln(P_t/P_{t-1})$. This measures appreciation/depreciation for the foreign exchange rates and returns for the stock markets.

The possible interconnection between the two types of markets is explored using the conditional correlation coefficients. To estimate the conditional correlation coefficients, this study employs the widely used and flexible Diagonal Vector GARCH (VECH) version of the MVGARCH (multivariate generalized autoregressive conditional heteroscedasticity) model (Bollerslev et al. 1988). This model is frequently used in modelling financial time series on the presumption that the variance–covariance matrix of financial market returns vary over time. Following Bollerslev et al. (1988), the model could be described as

$$\mathrm{VECH}(CV_t) = A + B \cdot \mathrm{VECH}\left(E_{t-1.}E'_{t-1}\right) + C \cdot \mathrm{VECH}(CV_{t-1}) \tag{1}$$

$$E_t|\psi_{t-1} \sim N(0, CV_t)$$

CV_t is $n \times n$ conditional variance–covariance matrix. E_t is 2×1 innovation vector, ψ_{t-1} is the information set at time $t - 1$, A is $N(N + 1)/2 \times 1$ parameter vector and C and B are $N(N + 1)/2 \times N(N + 1)/2$ parameter matrices. Since the number of parameters to be estimated might be a problematic issue, Bollerslev et al. (1988) as well as Goeij and Marquering (2004) suggested using the diagonal form of C and B.

The estimated conditional correlation coefficients are shown in Figs. 3 and 4.

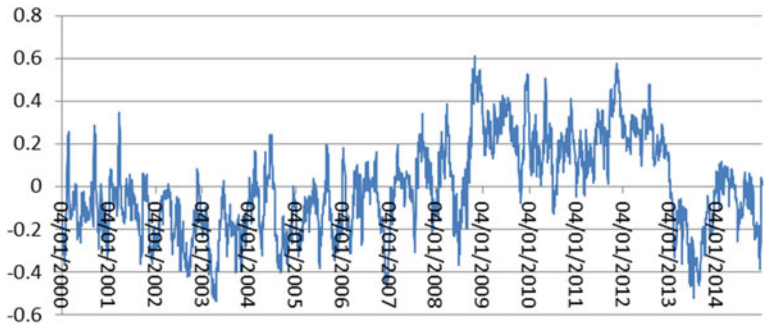


Fig. 3 Conditional correlation between FTSE and GBP/USD

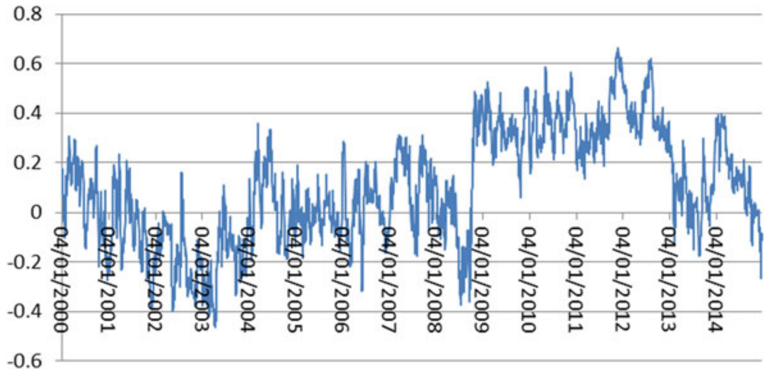


Fig. 4 Conditional correlation between DJIA and GBP/USD

The UK stock market and the GBP/USD market were mostly negatively correlated before the crisis of 2008. Over the period of crisis and during the aftermath of it, the two financial markets became positively correlated. The maximum value of correlation coefficient stood at 0.6. Since 2003 onwards, once again the markets have come to be negatively correlated. Figure 4 shows the conditional correlation between the US market and the GBP/USD rate.

On the contrary, the US stock market and the foreign exchange market have mostly been positively correlated, except for a period from 2002 to 2004. During the period of crisis of 2008, the extent of correlation increased, but it was never more than 0.6. In recent years, the correlation coefficient values have fallen to some extent. They, however, have remained positive.

Thus, the two markets are related differently to the corresponding foreign exchange market. While the UK stock market is mostly negatively correlated, the US market is positively correlated with the foreign exchange market. The period of crisis, however, is characterized by a positive correlation in both the cases. Moreover, the correlation values have never been excessively high. For the non-crisis period, it ranged mostly between –0.4 and 0.3 for both the stock market-foreign exchange market pairs.

This presence of correlation, however, says hardly anything about the predictability of the foreign exchange market crisis from the other financial markets. The study now explores into the possible causal relationship between the two types of markets using the method of Granger causality testing. Although the available literature does not concede the Granger causality testing as a true test for causality, it might give us some idea regarding the possible "lead-lag" relationship between the variables concerned.

The Granger (1969) approach to the question of whether x causes y is to see how much of the current y can be explained by past values of y and then to see whether adding lagged values of x can improve the explanation. y is said to be Granger-caused by x if y helps in the prediction of x, or equivalently if the coefficients on the lagged x's are statistically significant. However, the fact that x Granger causes y does not imply that y is the effect or the result of x. Granger causality measures precedence and information content but does not by itself indicate causality in the more common use of the term. Specifically, we run a bivariate model where

$$\left.\begin{aligned} Y_t &= a_0 + a_1 \cdot Y_{t-1} + \ldots + a_l \cdot Y_{t-1} + b_1 \cdot X_{t-1} + \ldots + b_l \cdot X_{t-1} + e_t \\ X_t &= a_0 + a_1 \cdot X_{t-1} + \ldots + a_l \cdot X_{t-1} + b_1 \cdot Y_{t-1} + \ldots + b_l \cdot Y_{t-1} + u_t \end{aligned}\right\} \tag{2}$$

And, we test the hypothesis: H_0: $b_1 = b_2 = \cdots = b_l = 0$.

The study tests for causality between financial markets for the three sub-phases: pre-2008 crisis, crisis period of 2008 and the aftermath of the crisis. The results are summarized in Table 1.

In all the cases, the US stock market leads the GBP/USD rate and the UK stock market. The UK market, however, leads neither the foreign exchange market nor the US stock market.

This, perhaps gives us a hint that crisis in the GBP/USD foreign exchange market perhaps might be predicted from the crises in the US stock market. To explore it in detail the study now introduces the stress index.

Table 1 Granger causality test—FTSE 100, DJIA and GBP/USD

Null hypothesis	Probability of accepting H_0			
	2000–2014	2000–2008	2008–2009	2009–2014
FTSE does not Granger Cause GBP_USD	0.29	0.15	0.17	0.19
GBP_USD does not Granger Cause FTSE	0.85	0.32	0.31	0.29
DJ does not Granger Cause GBP_USD	0.00	0.00	0.00	0.00
GBP_USD does not Granger Cause DJ	0.18	0.56	0.48	0.42
DJ does not Granger Cause FTSE	0.00	0.00	0.00	0.00
FTSE does not Granger Cause DJ	0.00	0.48	0.01	0.52

Table 2 Granger causality test—stress indexes for FTSE 100, DJIA and GBP/USD

Null hypothesis	Probability of accepting H_0			
	T = 15	T = 30	T = 45	T = 60
Stress in FTSE does not Granger Cause Stress in GBP_USD	0.41	0.56	0.34	0.36
Stress in GBP_USD does not Granger Cause Stress in FTSE	0.61	0.54	0.49	0.47
Stress in DJ does not Granger Cause Stress in GBP_USD	0.38	0.54	0.47	0.51
Stress in GBP_USD does not Granger Cause Stress in DJ	0.65	0.80	0.75	0.79
Stress in DJ does not Granger Cause Stress in FTSE	0.05	0.00	0.00	0.00
Stress in FTSE does not Granger Cause Stress in DJ	0.23	0.96	0.97	0.98

Stress indexes for financial markets are available in current literature. This study follows the approach of Patel and Sarkar (1998) and of Vila (2000) with some modification to identify crises in the context of the constructed portfolios. This approach has been followed by Chakrabarti and Sen (2014). Patel and Sarkar (1998) introduced a method called "CMAX method", which is a hybrid volatility loss measure. In the method the stress index is constructed as follows:

$$\mathrm{CMAX} = X_t / \max[X \in (X_{t-j} | j = 0, 1, \ldots, T]$$

where X_t is the financial market index. The moving window is determined by T. Hence, CMAX compares the current value of a variable with its maximum value over the previous T periods. Vila (2000) used this method to identify periods of slide in the stock market. The trigger level is considered at either 1.5 or 2 standard deviations below the mean of the series.

In this study, a stress index similar to CMAX is defined for four windows. Specifically it selects T to be equal to 15, 30, 45 and 60. Hence, it compares the current value of the financial variable with the maximum value over the previous 15, 30, 45 and 60 days. Thus, stress is defined for four periods of different lengths. A particular market is in stress, if stress index is less than 2 standard deviations below the mean of the return for that market. In that case, the current return of the market falls significantly below the historical market return. The study now inquires whether and how the stress in the foreign exchange market is 'caused' by stresses in the associated stock markets. The Granger causality results are shown in Table 2.

The stress in GBP/USD currency rate is not caused by stresses in any of the associated stock markets. The stress in the US stock market, however, causes stress in the UK stock market. But for the developed market currency pairs, stress in other financial markets cannot lead or lag stress in the foreign exchange market.

The study now explores the same relationship for the two other cases namely the case for a developed and an emerging market (that is the India–US market pair) and the case for two emerging markets (that is the India–Singapore market pair).

3.2 Stress in the INR/USD Exchange Rate and the Two Stock Markets

The study explores the relationship between the INR/USD exchange rate and the two stock indexes (namely, the BSE SENSEX and DJIA) following the same methodology that was followed in the earlier section. The movements in the INR/USD exchange rate and the SENSEX and DJIA are shown in Figs. 5 and 6.

The foreign exchange market faced a crisis in 2008. It is characterized by no other significant crisis over the 15-year period. The stock market movements are shown in Fig. 6.

Both the stock markets faced crisis during 2008. Before the crisis, the DJIA outperformed the SENSEX, while in the post-crisis period SENSEX has performed

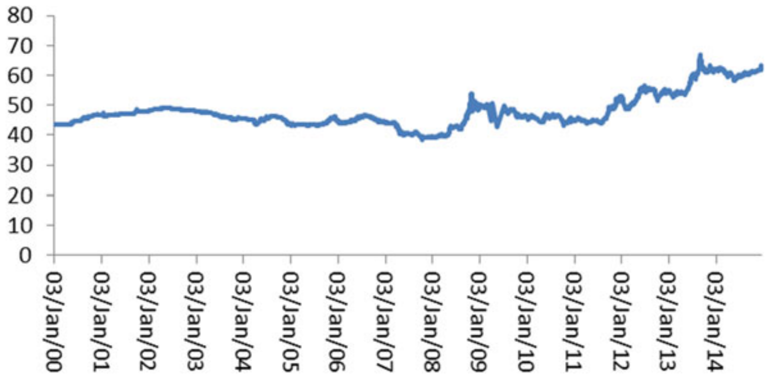


Fig. 5 Movements in INR/USD series (2000–2014)

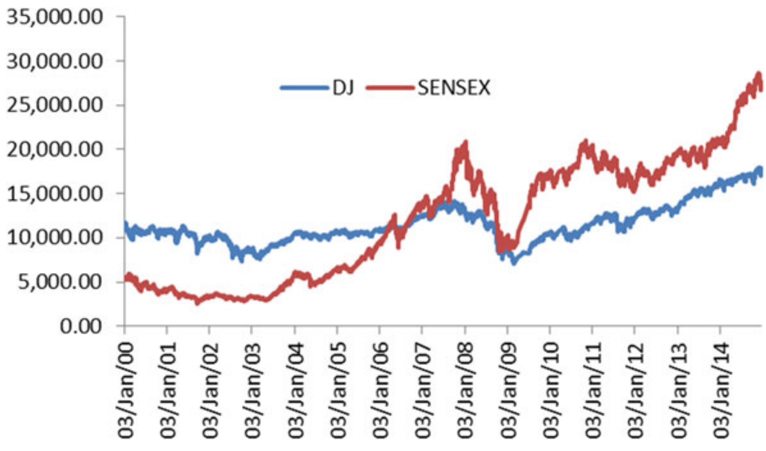


Fig. 6 Movements in SENSEX and DJIA series (2000–2014)

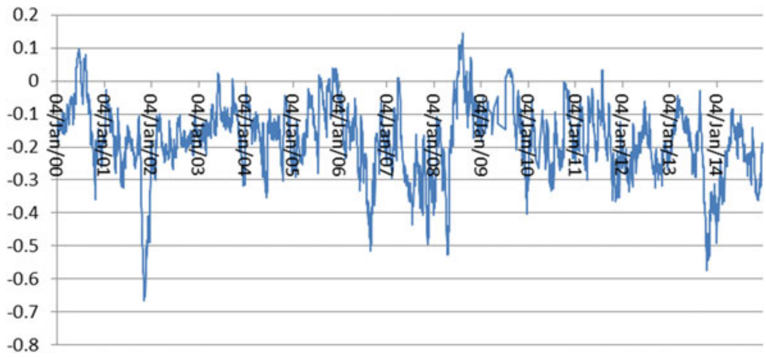


Fig. 7 Conditional correlation between SENSEX and INR/USD

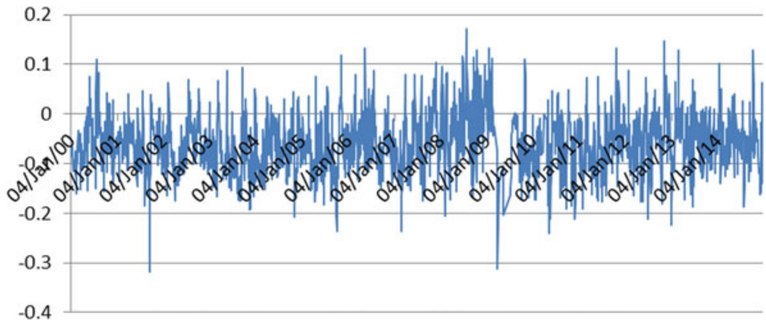


Fig. 8 Conditional correlation between DJIA and INR/USD

better than DJIA. The movements across the financial markets are once again similar but the association among them is yet to be explored.

Figures 7 and 8 depict the conditional correlation between the two financial market pairs, namely the INR/USD-SENSEX and the INR/USD-DJIA pairs.

The conditional correlation between SENSEX and INR/USD has been negative except for two or three extreme cases. Moreover, the correlation values ranged between 0 and –0.3 in most cases. In only a few cases the coefficient fell to –0.5. Over the crisis period, the correlation became positive only for a very small span. The value, however, was very low. Figure 8 shows the conditional correlation between DJIA and INR/USD over time.

The movement in the conditional correlation between DJIA and INR/USD has been quite different from the other cases considered earlier in the study. The conditional correlation fluctuated but remained within the band of 0.1 to –0.1. The crisis period witnessed no significant change in the nature of the movement in the conditional correlation.

Table 3 Granger causality test—SENSEX, DJIA and INR/USD

Null hypothesis	Probability of accepting H_0			
	2000–2014	2000–2008	2008–2009	2009–2014
SENSEX does not Granger Cause INR_USD	0.00	0.00	0.00	0.00
INR_USD does not Granger Cause SENSEX	0.00	0.02	0.39	0.00
DJ does not Granger Cause INR_USD	0.00	0.00	0.00	0.00
INR_USD does not Granger Cause DJ	0.53	0.86	0.51	0.84
DJ does not Granger Cause SENSEX	0.00	0.00	0.00	0.00
SENSEX does not Granger Cause DJ	0.61	0.18	0.53	0.16

Table 4 Granger causality test—stress indexes for SENSEX, DJIA and INR/USD

Null hypothesis	Probability of accepting H_0			
	T = 15	T = 30	T = 45	T = 60
Stress in SENSEX does not Granger Cause Stress in INR_USD	0.00	0.00	0.00	0.00
Stress in INR_USD does not Granger Cause Stress in SENSEX	0.06	0.06	0.06	0.06
Stress in DJ does not Granger Cause Stress in INR_USD	0.00	0.00	0.00	0.00
Stress in INR_USD does not Granger Cause Stress in DJ	0.71	0.70	0.77	0.80
Stress in DJ does not Granger Cause Stress in SENSEX	0.00	0.00	0.00	0.00
Stress in SENSEX does not Granger Cause Stress in DJ	0.01	0.01	0.01	0.02

The nature of movement in the conditional correlation, however, does not imply any causality between the occurrences of crises in the two types of markets. The possible causality between the two markets is then explored employing the Granger causality test for the market pairs. The results are summarized in Table 3.

The causality results obtained for the developed-emerging market pair is different from those obtained for those for the developed-developed market pair. While the developed-country stock market was leading the stock market of the emerging economy, the individual stock markets were leading the corresponding foreign exchange rate in all the three sub-phases. Except for the crisis period, the emerging-economy stock market has been related with the foreign exchange market with bi-way causality.

Table 4 shows the results obtained for Granger causality testing among the stress indexes for the financial markets.

The results obtained for the developed-emerging market pair is once again different from those obtained for the developed-developed market pair. For all the chosen windows, the stresses in the US and the Indian market were causing stresses in the INR/USD exchange rate. The stock market stress indexes, however, are related by a both-way causality.

It now remains to explore the relationship for the emerging-emerging market pair and that is where the study moves next.

Fig. 9 Movements in SGD/INR series (2000–2014)

Fig. 10 Movements in SENSEX and STI series (2000–2014)

3.3 Stress in the INR/SGD Exchange Rate and the Two Stock Markets

Figures 9 and 10 depict the movements in the BSE SENSEX, STI and SGD/INR over time.

The foreign exchange market witnessed a dip in 2008. However, the crisis was more acute in the context of developed-developed market pair. Since 2009, the SGD/INR rate has increased considerably. Like the two other currency pairs, SGD/INR has not experienced any other major crisis.

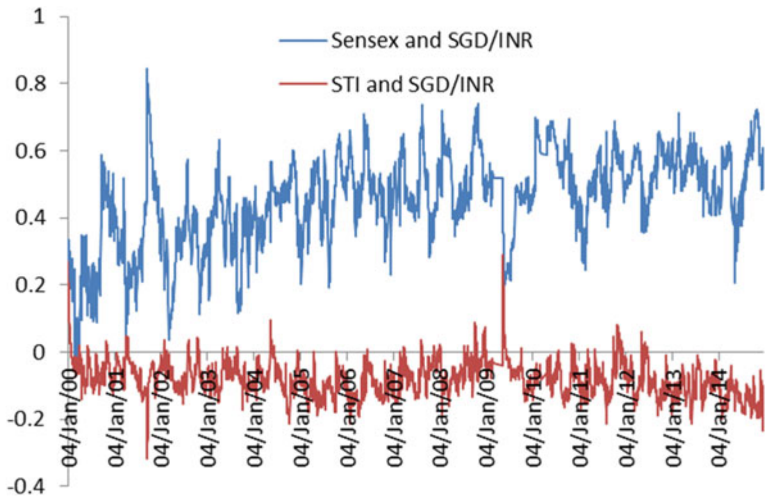


Fig. 11 Conditional correlation between market pairs

Table 5 Granger causality test—SENSEX, STI and SGD/INR

Null hypothesis	Probability of accepting H_0			
	2000–2014	2000–2008	2008–2009	2009–2014
SENSEX does not Granger Cause SGD/INR	0.00	0.00	0.00	0.00
SGD/INR does not Granger Cause SENSEX	0.94	0.92	0.85	0.83
STI does not Granger Cause SGD/INR	0.00	0.00	0.00	0.00
SGD/INR does not Granger Cause STI	0.19	0.18	0.16	0.14
STI does not Granger Cause SENSEX	0.02	0.00	0.00	0.00
SENSEX does not Granger Cause STI	0.00	0.00	0.00	0.00

Figure 10 depicts the movements in the two emerging stock markets.

Just like the Indian stock market, the Singapore market experienced a crisis in 2008. The market has recovered since then.

Figure 11 shows the conditional correlation between the two market pairs.

The conditional correlation between STI and SGD/INR has always been negative and remained within the band of 0–0.2. The conditional correlation between SENSEX and SGD/INR has always been positive and remained within the band of 0.1–0.7. The possible presence of causality is reported in Table 5.

The two stock markets are related by both-way causality, where as the individual stock markets lead the foreign exchange market. The tests for possible causality, however, reveal significantly different results. For all the stress windows chosen, stresses in the Indian stock market lead the stress in the SGD/INR. There is no other causal relationship present in the stress indexes (Table 6).

Table 6 Granger causality test—stress indexes for SENSEX, STI and SGD/INR

Null hypothesis	Probability of accepting H_0			
	T = 15	T = 30	T = 45	T = 60
Stress in SENSEX does not Granger Cause Stress in SGD/INR	0.01	0.01	0.01	0.01
Stress in SGD/INR does not Granger Cause Stress in SENSEX	0.83	0.86	0.82	0.82
Stress in STI does not Granger Cause Stress in SGD/INR	0.24	0.23	0.23	0.25
Stress in SGD/INR does not Granger Cause Stress in STI	0.31	0.28	0.31	0.33
Stress in STI does not Granger Cause Stress in SENSEX	0.27	0.23	0.28	0.29
Stress in SENSEX does not Granger Cause Stress in STI	0.16	0.16	0.17	0.16

4 Concluding Remarks

The present study has been an exploration into the possibility of predicting stress in foreign exchange market from that in other financial markets, particularly the stock markets. The study however, differs from the approaches available in the existing literature in the sense that instead of delving into the mere association and causality among market returns it explores the presence of possible causality among stresses in different markets. In the process, it defines a stress index for each market. Over a period of 15 years ranging from 2000 to 2014, the study finds stress in foreign exchange market to be ‘caused’ by those in stock markets. The relationship, however, is not always independent of the level of financial development of the markets. For the two developed markets, the stock markets and the corresponding foreign exchange market were positively correlated during the period of crisis. But in the non-crisis periods, the correlation was negative for the UK market and positive for the US market. The situation is similar for the two emerging market pairs. While the correlation between the Indian stock market and SGD/INR pair has always been positive, it has been negative for the STI and the SGD/INR pair. The observed results are completely different for the developed-emerging market pairs. The conditional correlation between SENSEX and INR/USD has been negative except for two or three extreme cases, while it fluctuated within a band of 0.1–0.1 for the US market. For the developed-developed market pair, however, stress in foreign exchange market is not at all caused by the corresponding developed-country stock markets. For the US–India pair, stress in INR/USD rate is caused by stresses in both the stock markets. For the emerging-emerging market pair, stresses in Indian stock market causes the SGD/INR exchange rate but not the other way round. Thus, for the developed market, the stresses in the stock market are not causing (and are not even caused by) stresses in the foreign exchange market. However, for the emerging markets, particularly for India, the channel of stress transmission remains open between the two stock markets and the foreign exchange market. Hence, foreign exchange market crisis and stock market crisis (whether generated domestically or emanating from the developed, foreign market) may appear as “twin”. The case for Singapore, however, is slightly different from that of India. There remains a channel of stress transmission from the stock market to the foreign exchange market, where stress is generated only in the other emerging market. Thus, the emerging markets that experience huge inflow of foreign capital in their stock markets might take stock market crises and foreign exchange market crises as twins. The policy implications, however, might differ. In some cases, it would be enough to regulate the domestic stock market, but in some other instances crises may be contagious coming from stock markets abroad.

References

Ajayi RA, Friedman J, Mehdian SM (1998) On the relationship between stock returns and exchange rates: test of granger causality. Global Finance J 9(2):241–251

Bahmani-Oskooee M, Sohrabian A (1992) Stock prices and the effective exchange rate of the dollar. Appl Econ Taylor Francis J 24(4):459–464

Bollerslev T, Engle RF, Wooldridge JM (1988) A Capital Asset Pricing Model with time varying covariances. J Polit Econ 96:116–131

Calvo G (1995) Varieties of capital market crises. University of Maryland, Mimeo

Chakrabarti G, Sen C (2014) Green investing: a case for India. Springer, Berlin

Choi DFS, Fang V, Fu TY (2010) Volatility spillovers between stock market returns and exchange rate changes: the New Zealand case. http://www.mssanz.org.au/MODSIM07/papers/40_s2/VolatilitySpilloverss2Choi.pdf. Accessed 9 Mar 2014

Goeij de PC, Marquering W (2004) Modeling the conditional covariance between stock and bond returns: A multivariate GARCH approach. J Financ E 2(1):531–564

Granger CWJ (1969) Investigatine causal relations by econometric models and cross-spectral methods. Econometrica 37(3):424–438

Granger CJ, Huang B, Yang C (2000) A bivariate causality between stock prices and exchan ge rates: evidence from recent Asian flu. Q Rev Econ Finance 40:337–354

Horobet A, Livia I (2007) On the dynamic link between stock prices and exchange rates: evidence from Romania. http://mpra.ub.uni-muenchen.de/6429/1/MPRA_paper_6429.pdf. Accessed 11 Dec 2013

Kaminsky G, Reinhart C (1999) The twin crises: the causes of banking and balance of payments problems. Am Econ Rev 89(3):473–500

Kasman S (2005) The relationship betweem exchange rates and stock prices: a causality analysis. J Soc Sci Inst 5:70–79

Mishkin F (1996) Understanding financial crises: a developing country perspective. Annual world bank conference on development economics. World Bank, Washington, pp 29–62

Mishra AK (2004) Stock market and foreign exchange market in india: are they related? South Asia Econ J 5:209–232

Muhammad N, Rasheed A (2002) Stock prices and exchange rates: are they related? evidence from South Asian countries. Pak Dev Rev 41(4):535–550

Nath GC, Samanta GP (2003) Relationship between exchange rate and stock prices in India—an empirical analysis. http://www.papers.ssrn.com/sol3/papers.cfm?abstract_id=475823. Accessed 30 June 2014

Pan M, Fok RC, Liu YA (2007) Dynamic linkages between exchange rates and stock prices: evidence from East Asian markets. Int Rev Econ Finance 16:503–520

Patel, S. and A. Sarkar, (1998), "Crises in Developed and Emerging Stock Markets", Financial Analysts Journal (November/December), pp. 50–61

Reinhart CM, Vegh C (1996) Do exchange rate-based inflation stabilizations sow the seeds of their own destruction?. University of Maryland and UCLA, Mimeo

Roubini N, Mihm S (2010) Crisis Economics. Allen Lane, London

Shew JA Jr (2008) Causality relationship between foreign exchange rates and stock market close: evidence in Singapore. Bryant economic research paper, 1(11). http://www.bryant.edu/wps/wcmresources/libfiles/economics/research/Vol1No11.pdf. Accessed 20 Oct 2014

Shiller RJ (2003) From efficient market theories to behavioural finance. J Econ Perspect 17 (2003):90

Stavárek D (2005) Stock prices and exchange rates in the EU and the United States: evidence on their mutual interactions. Czech J Econ Finance (Finance aUver) 55(3–4):141–161

Stoker J (1995) Intermediation and the business cycle under a specie standard: the role of the Gold Standard in English financial crises, 1790–1850. University of Chicago, Mimeo

Tabak BM (2006) The dynamic relationship between stock prices and exchange rates: evidence for Brazil. Int J Theor Appl Finance 9(08):1377–1396

Velasco A (1987) Financial and balance of payments crises. J Dev Econ 27:263–293

Vila A (2000) Asset price crises and banking crises: some empirical evidence. BIS Conference Papers 8(March):232–252

Wickremasinghe GB (2006) Dynamic relations between stock prices and exchange rates in Sri Lanka: some empirical evidence. In: 11th Finsia-Melbourne centre for financial studies banking and finance conference, 25th September 2006, RMIT University, 239 Bourke Street, Melbourne, Australia. http://vuir.vu.edu.au/1295/. Accessed 12 Feb 2009

Yang SY, Doong SC (2004) Price and volatility spillovers between stock prices and exchange rates: empirical evidence from The G-7 countries. Int J Bus Econ 3(2):139–153

The Impossible Trinity: Where Does India Stand?

Rajeswari Sengupta

Abstract The Global Financial Crisis of 2008 and the heightened macroeconomic and financial volatility that followed the crisis raised important questions about the current international financial architecture as well as about individual countries' external macroeconomic policies. Policy-makers dealing with the global crisis have been confronted with the 'impossible trinity' or the 'Trilemma,' a potent paradigm of open economy macroeconomics asserting that a country may not target the exchange rate, conduct an independent monetary policy and have full financial integration, all at the same time. This issue is highly pertinent for India. A number of challenges have emanated from India's greater integration with the global financial markets during the last two decades, one of which includes managing the policy tradeoffs under the Trilemma. In this chapter, I present a comprehensive overview of a few empirical studies that have explored the issue of Trilemma in the Indian context. Based on these studies I attempt to analyze how have Indian policy-makers dealt with the various trade-offs while managing the Trilemma over the last two decades.

Keywords Impossible trinity · Financial integration · Currency stabilization · International reserves · Sterilized intervention

JEL Classification F3 · F4 · F6

Views expressed in this chapter are entirely my own and do not reflect that of my co-authors with whom I have written the papers that have been surveyed here.

R. Sengupta (✉)
Indira Gandhi Institute of Development Research (IGIDR), Gen. A. K. Vaidya Marg, Goregaon East, Mumbai 400065, India
e-mail: rajeswari.sen@gmail.com

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_25

1 Introduction

The Global Financial Crisis (GFC) of 2008 and the heightened macroeconomic and financial volatility that followed the crisis raised important questions about the current international financial architecture as well as about individual countries' external macroeconomic policies. Policy-makers dealing with the global crisis have been confronted with the 'impossible trinity' or the Trilemma, a crucial paradigm of open economy macroeconomics asserting that a country may not be able to target the exchange rate, conduct an independent monetary policy and have full financial integration, all at the same time. The fast-paced and massive financial globalization undertaken by most emerging economies over the past 20 years along with the rapid deepening of domestic and international financial markets, have significantly modified the context of the Trilemma paradigm.

The 'impossible trinity' or the Trilemma is a fundamental contribution of the Mundell–Fleming framework. It states that policy makers in open economies often face a three-way trade-off between the contradictory policy objectives namely, a stable exchange rate, an independent monetary policy designed to target domestic inflation and open capital markets. Confronted with this Trilemma policy makers have to choose a combination of any two objectives because all three goals can be mutually inconsistent. For instance, supposing policy makers in an economy with fixed exchange rate raise the interest rate in order to control domestic inflation—a move that also increases the difference with world interest rate. Given arbitrage in open capital markets and the underlying interest rate parity condition, such a move will attract capital inflows into the domestic economy resulting in an appreciating pressure on the fixed exchange rate. Subsequent intervention in the foreign exchange market to buy international reserves and sell domestic currency in order to maintain the fixed exchange rate in the face of the appreciation pressure will end up defeating the original objective of controlling domestic inflation. Thus, while the US is an example of an economy pursuing flexible exchange rates, open capital markets and independent monetary policy, countries belonging to the Euro zone follow a combination of exchange rate stability, financial integration, and zero monetary independence (Aizenman 2010). In recent times, empirical estimation of the Trilemma has been popularized in the academic literature most notably by Obstfeld et al. (2005, 2010), Aizenman et al. (2008, 2010a, b, 2011).

The issue of the Trilemma is highly pertinent for India—a major emerging economy in a world where emerging or transitioning countries are increasingly playing a crucial role in restoring global economic growth in the post-crisis era. A number of challenges have emanated from India's greater integration with the global financial markets during the last two decades, one of which includes managing the policy tradeoffs presented by the Trilemma. The importance of the Trilemma in present day macroeconomic context in India cannot be overstated. Yet there have been very few empirical studies that have investigated this issue for India. This is perhaps because for the longest of time India has followed a pegged exchange rate regime, and a relatively closed capital account compared to other emerging economies. However, with the passage of time and increasing integration with the global economy, a shift away from pegged exchange rate regime and steady albeit cautious liberalization of the capital account have made the macroeconomic management of the Trilemma policy objectives increasingly complex.[1]

Halfway into the current decade, India is experiencing several interesting phenomena in its macroeconomic environment. Its GDP growth rate is increasing once again after few years of slowdown in the aftermath of the Global Financial Crisis; its equity and debt markets continue to attract foreign capital inflows buoyed by the economy's revitalized growth prospects and improving macroeconomic scenario; it is on the verge of embarking on a 'glide-path' of inflation targeting following the release of the Urjit Patel Committee Report by the RBI in 2014, and finally, it continues to face growing currency volatility owing to a rapidly changing post-crisis global environment. In view of these, the question of where India stands with regard to the Trilemma becomes all the more pertinent now.

In this chapter, I present a comprehensive overview of a select few empirical papers that I have written with other co-authors, investigating the evolution of the Trilemma in India over the past couple of decades and analyzing the extent of tradeoffs faced by policy makers between financial integration, monetary autonomy and exchange rate stability. Based on these studies I attempt to analyze how have Indian policy makers dealt with the various trade-offs while managing the Trilemma over the last two decades and draw relevant policy conclusions.

2 India Post 1991 Liberalization Reforms

After the Balance of Payments crisis in 1991, a comprehensive series of liberalization, privatization and deregulation reforms were implemented in the banking sector, trade sector as well as financial markets in India, the net result being, over the next couple of decades the Indian economy witnessed several structural changes (Shah 2008; Mohan and Kapur 2009; Hutchison et al. 2011). In what follows, I provide a brief account of the Indian economy with respect to the two main tenets of the Trilemma-capital account openness and exchange rate stability.

2.1 Capital Account Management

India's capital account liberalization that began in early 1990s has since then been a steady albeit slow and gradual process (Hutchison et al. 2012; Sen Gupta and Sengupta 2014). As regards opening up the financial markets to foreign investment, Indian authorities have always proceeded with a lot of caution and apparent

[1]In this chapter financial integration and capital account openness have been used interchangeably to mean the integration of the domestic economy with global financial markets.

skepticism about the benefits of foreign capital balanced against the potential instability that such capital inflows might trigger. While on one hand foreign capital offers diversification opportunities to investors and provides avenues to bridge the gap between domestic saving and investment, on the other hand unbridled flows can fuel inflationary pressures, fan asset price bubbles, and sudden reversals in capital inflows can lead to instabilities and even crises in financial and currency markets. The decade of the 1990s was replete with several such incidents of financial crises all over the emerging world as seen in Mexico (1994), East Asia (1997), Russia (1998), Brazil (1999), and eventually Argentina (2000).

All these episodes of crises were triggered by rampant volatile debt inflows into the domestic economies. It is not surprising therefore that when its turn came after the crisis of 1991 India adopted a calibrated and hierarchical approach toward capital account liberalization, prioritizing some flows over others (Reddy 2008). With the memories of the emerging economy crises fresh in mind, India gradually opened up non-debt flows (FDI, and portfolio) more than debt flows.

From early 2000s onwards India has been receiving significant amount of foreign investment. India experienced as many as three separate capital inflow surge episodes between 2000 and 2008 as analyzed by Sen Gupta and Sengupta (2015) using the methodology employed by Forbes (2014) to identify periods of capital surges and stops. This was the period when interest rates in the advanced world were relatively lower, emerging economies had started registering impressively high growth rates and their domestic macroeconomic fundamentals looked robust post recovery from the slew of financial crises of the 1990s, emerging economy central banks were also flush with foreign currency reserves and hence looked capable of defending any speculative attack on their respective currencies, and foreign investors in search of yields began investing massively in the equity and debt markets of the emerging world. India with an average growth rate of 9 % and interest rates much higher than that of the advanced world was needless to say an attractive destination for these investors seeking higher yields. Net capital flows increased from $17.3 billion in 2003–04 to over $107.9 billion in 2007–08.

The pre-GFC surge episodes were primarily driven by bank and non-bank flows such as commercial borrowings by the Indian corporate sector, short-term trade credits, and deposits by non-resident Indians as well as by portfolio equity flows. These inflows were encouraged by the widening interest rate differential between India and the advanced economies, greater liberalization of borrowing norms, improved domestic economy fundamentals and availability of abundant global liquidity. However, though in absolute terms India is now more open than it used to be during the early 1990s, the cautious and gradual approach toward capital account liberalization has meant that India has significantly lagged behind other countries irrespective of whether the capital account openness is measured using de-jure or de facto indicators as shown by Sen Gupta and Sengupta (2013). Even today, more than two decades after India started opening up its capital account, there exists an extensive array of restrictive controls imposed by authorities on different categories of foreign investment in order to actively manage the capital account.

2.2 Exchange Rate Management

India officially moved to a market-based exchange rate system in 1993Q3. Empirical analysis of the data shows that the nominal exchange rate has gone through several structural breaks since then. Zeileis et al. (2010), Patnaik et al. (2011) show that while the Rupee was closely pegged to the dollar till 1997, during the Asian financial crisis of 1997–98, there was a short period of flexibility when the Rupee depreciated significantly.[2] Thereafter from 1998 to 2004 the Rupee was once again tightly pegged to the Dollar. 2004 onwards the peg to the Dollar was replaced by a basket peg to several other international reserve currencies such as Euro, Yen, etc., implying that the exchange rate exhibited greater flexibility than before. The basket peg sort of ended around 2007 and gave way to a more flexible exchange rate regime that continued till 2013 July. In the summer of 2013, when the US Federal Reserve chairman Ben Bernanke announced a possible tapering of its Quantitative Easing (QE) program, large amounts of foreign capital fled India to safety driven by risk aversion thereby resulting in a sharp depreciation of the Rupee. The RBI intervened in the foreign exchange market in order to stabilize the Rupee and reduce the currency volatility.

Excessive foreign exchange intervention unless sterilized can fuel inflationary pressures by increasing the reserve money base. So in order to insulate the domestic economy from rising inflation central banks intervening in foreign exchange market to prevent a currency appreciation, conduct contractionary open market operations by selling bonds so as to absorb the excess liquidity from the system. The Rupee appreciated significantly in response to the surge in capital inflows from 2000 onwards thereby adversely affecting the competitiveness of Indian exports. During this time the RBI intervened actively in the foreign exchange market to prevent the Rupee from appreciating, given that India has had a long history of current account deficit and a nominal appreciation would have further aggregated the deficit. Till 2004 RBI sterilized this intervention by selling government bonds, raising the reserve requirements of the commercial banks by the concomitant amount and thereby reducing liquidity in the system.

RBI continued the sterilization till 2003 when it ran out of government bonds to sell. The Market Stabilization Scheme (MSS)—a new instrument for sterilization, was started around this time and RBI began issuing MSS bonds to sterilize its intervention but the rising fiscal costs of interest payment on bonds meant that sterilization could only be partial. This timing also happened to coincide with increased flexibility in the Indian rupee vis-à-vis the US dollar. Sen Gupta and Sengupta (2013) also find that RBI's intervention in the foreign exchange market has been sort of asymmetric, intervening when the exchange rate has been appreciating and adopting a hands off approach during times of depreciation, either for fear of losing reserves or to let a depreciating currency ameliorate a growing current account deficit.

---

[2]Here Dollar is used to imply the US dollar.

The above discussion throws some light on where India stands with regard to capital account as well as exchange rate management. It seems India's capital account is only partially open and the exchange rate is in a managed floating regime. India has also been accumulating massive amounts of foreign currency reserves since the early 2000s. Foreign exchange reserves climbed from around \$150 billion in mid-2005 to over \$300 billion in mid-2010, a doubling in just 5 years and making India one of the largest reserve-holding countries in the world (Sen Gupta and Sengupta 2013).

3 Impossible Trinity or Trilemma

The pioneering work in empirically assessing Trilemma as mentioned earlier, has been done among others, by Aizenman et al. (2010a, b, 2011), henceforth referred to as ACI. They measured the Trilemma indices, namely monetary independence (MI), exchange rate stability (ERS) and capital account openness (KO) or financial integration over a large cross-section of countries—developed as well as developing. They also simultaneously incorporated the role played by foreign exchange reserves in macroeconomic management, which was quite a novel aspect, and hitherto not explored by others. ACI view reserves as a fourth dimension of these policy-tradeoffs. They find that most emerging economies operate in an intermediate range of partial financial integration and managed floating regimes with their respective central banks actively intervening in foreign exchange markets and using reserve management policies in the short-run to balance the trade-offs present by the Trilemma. Their seminal work in empirically measuring the Trilemma triggered a series of studies individually investigating evolution of the Trilemma for several emerging economies.

3.1 Evolution of Trilemma in India

Among the few empirical studies that have explored the issue of Trilemma in the Indian context, Hutchison et al. (2011) were perhaps one of the earliest. In fact, it was also the first paper that took the ACI panel-data approach in measuring the Trilemma and modified it suitably to estimate the Trilemma in a time-series framework for a single country. In their paper, they attempted to answer the questions: To what extent has financial integration imposed greater constraints on exchange rate and interest rate policies in India? What has been the cost of international financial liberalization in terms of macroeconomic policy?

They address these questions by measuring the trade-off between capital account openness (KO), exchange rate stability (ERS) and monetary independence (MI) in India. They calculate a Trilemma index for India and investigate its evolution over time. They also assess the impact of the Trilemma configurations on major macroeconomic indicators for India.

They use quarterly data from 1996Q2 to 2009Q3 and therefore can exploit greater data variation unlike ACI who had used annual data. They also divide the entire sample period into three equal segments to better study the evolution of Trilemma over time. They measure MI as the reciprocal of the correlation between interest rates in the home country (here India) and the base country (here the United States). They calculate quarterly correlations using weekly interest rate data on 90-day government securities. More precisely, MI is measured as:

$$\mathrm{MI} = \frac{1 - \mathrm{corr}\left(i^{\mathrm{Ind}}, i^{\mathrm{US}}\right)}{1 - (-1)} \tag{1}$$

This index can theoretically take any value ranging from 0 to 1 with a higher value indicating greater degree of monetary independence. The authors measure ERS using quarterly standard deviations of the change in the log of Rupee-Dollar exchange rate and construct the index based on the following formula:

$$\mathrm{ERS} = \frac{0.01}{0.01 + \mathrm{stdev}(\Delta(\log(\mathrm{exch\_rate})))} \tag{2}$$

Appropriate scaling ensures that the index lies between 0 and 1, with the highest value indicating the greatest degree of exchange rate stability.

Finally, to measure financial integration Hutchison et al. (2011) deviate significantly from ACI who use the de-jure Chinn-Ito index (Chinn and Ito 2008), which for India exhibits little or no variation over time and hence might not be a suitable measure of capital account openness in the context of the Trilemma. Hutchison et al. (2011) use a de facto measure of capital account openness based on the ratio of the sum of inward and outward foreign investment flows to GDP (Lane and Milessi-Ferreti 2007). Moreover, they use gross instead of net capital flows, in order to be able to distinguish between negligible flows and large flows. As pointed out succinctly by Sen Gupta and Sengupta (2013, 2014) as well, a country with high *de jure* openness can have low capital flows owing to limited opportunities for economic returns and hence be able to simultaneously stabilize exchange rate and retain monetary autonomy. Alternatively, a country with low *de jure* openness can experience large flows owing to low enforcement of capital controls, and face a trade-off between ensuring MI and ERS. The evolution of the three Trilemma indices during the sample period considered by Hutchison et al. (2011) is presented in the three panels in Fig. 1.

As mentioned before, an important contribution of ACI is connecting international reserves with the three Trilemma policy variables, the general idea being that reserve accumulation gives policymakers more flexibility in dealing with the short-run trade-offs between MI and ERS, for a given level of financial integration. Like other emerging economies, this is highly pertinent for India as well given the remarkable pace of reserve hoarding that the RBI has been undertaking especially

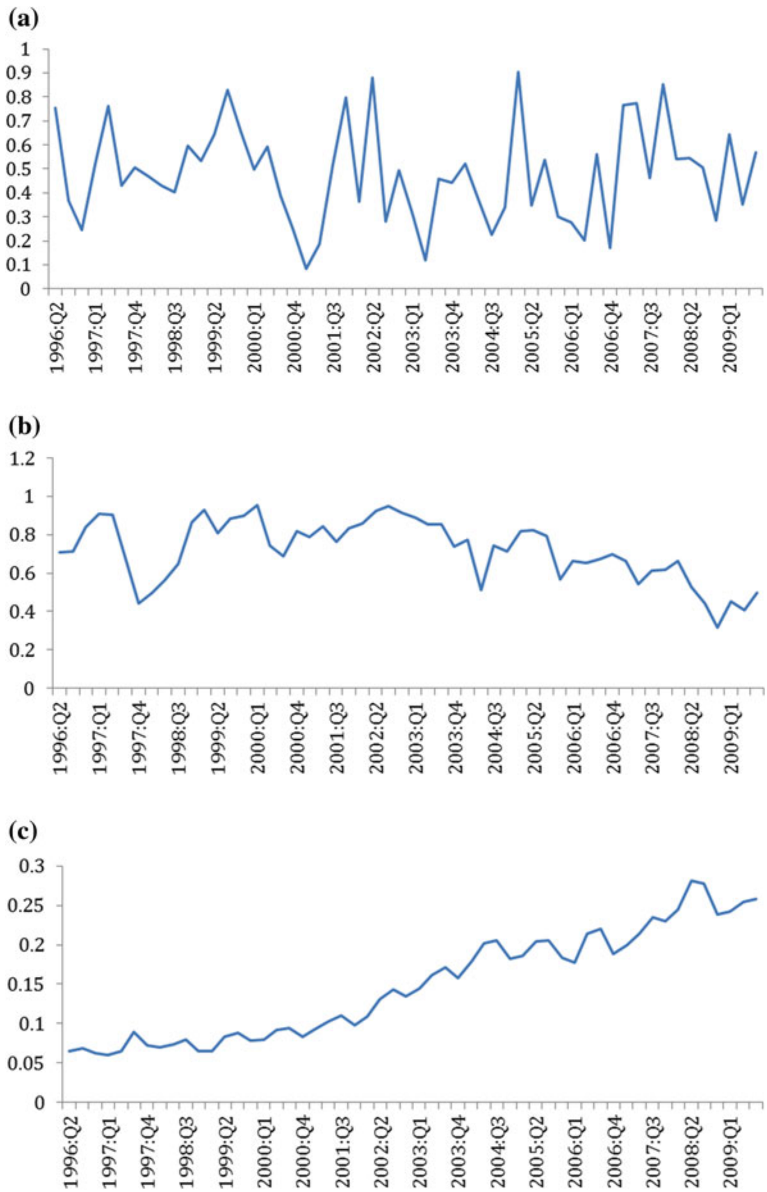


Fig. 1 **a** Monetary independence index. **b** Exchange rate stability index. **c** Capital account openness index. *Source* Hutchison et al. (2011)

since the early 2000s. In this context, Hutchison et al. (2011) demonstrate the evolution of the Trilemma policy variables as well as of the reserves-GDP ratio for India over the three windows spanning across their sample period, using the classic

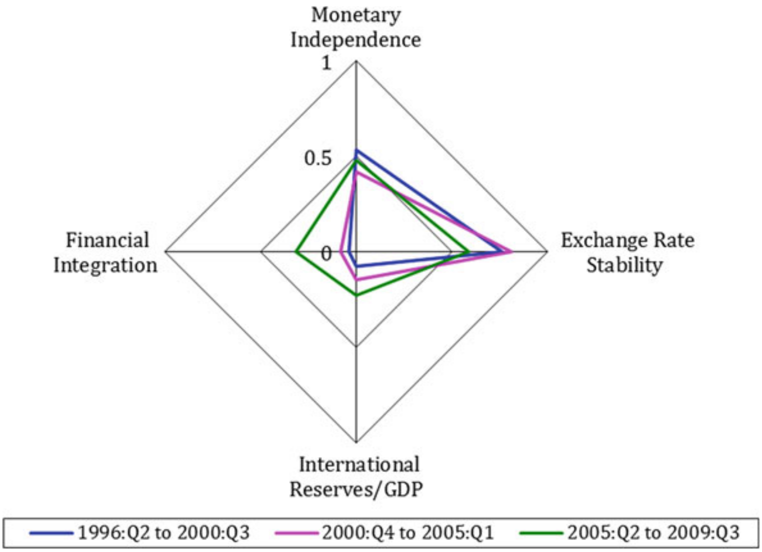


Fig. 2 The trilemma and reserve accumulation. *Source* Hutchison et al. (2011)

diamond graph popularized by ACI. In the diamond graph, the authors plot the averages of each of three Trilemma indices along with the reserves-GDP ratio for each subperiod. The diamond graph for India as seen in Fig. 2 shows that the Indian economy increased its ratio of reserves to GDP along with greater financial integration, as it tried to balance MI and ERS.

While policy makers cannot achieve all three Trilemma objectives at the same time, they can always strive to achieve a combination of all three. With this principle in mind, the authors regress a constant (in their case, two) on the three indices and try to assess the relative weight placed by the Indian policy makers on each of the three policies. They find that ERS received high policy weight throughout their sample period. Between 2000Q4 and 2005Q1, as capital openness increased, MI was completely lost, whereas there was an attempt to retain, or even strengthen ERS. This was the phase of surge in capital inflows when the RBI was actively intervening in the foreign exchange market to prevent a nominal currency appreciation and stabilize the value of the Rupee vis-à-vis the Dollar. In the final subperiod, i.e., between 2005Q2 and 2009Q3, as capital openness continued to increase, some ERS was sacrificed to recover MI.

These findings match well with the facts presented in the previous section that post 2007 nominal exchange rate exhibited greater flexibility perhaps largely owing to inability of the RBI to fully sterilize its foreign exchange intervention. In general the authors find that as in other emerging countries, financial integration has increased markedly after the mid-2000s in India, which is also in line with the discussion in the previous section. The rise in capital account openness in turn, has come with corresponding limitations especially on MI. They also investigate the question how the change in the Trilemma configuration has influenced inflation and inflation volatility in India. The authors find that greater MI is empirically and systematically associated with lower inflation. On the other hand, greater ERS lowers inflation volatility, perhaps by stabilizing import and commodity prices. However, as India struggles to balance financial globalization with domestic monetary stability, capital account openness is found to be associated with higher inflation volatility.

In a later study empirically comparing the Trilemma evolution in China and India, Aizenman and Sengupta (2013) use a methodology similar to that of Hutchison et al. (2011) over the period 1990Q1–2010Q4, and find that while China's Trilemma configurations are unique relative to other emerging markets in the predominance of ERS as a tool for macroeconomic management attained at the expense of MI and financial integration, in contrast, the Trilemma configurations of India seem more in line with the choices made by other emerging countries. India, like other emerging economies, has been balancing all three policy objectives and attaining a somewhat middle-ground perhaps through changes in its reserves stock, achieving comparable levels of ERS and financial integration. The authors find that ERS and financial integration were given marginally more importance followed by MI. In their rigorous econometric analysis, all three policy objectives come out statistically significant for India with relatively higher weight being placed on the ERS index, thereby echoing the findings of Hutchison et al. (2011). Overtime the Trilemma configuration that has evolved in India is one of greater exchange rate stability and financial integration, combined with an attempt to retain monetary autonomy through active intervention in foreign exchange markets.

Finally, in their recent work (2013, 2014) explore the various nuances of India's capital account management since the early 2000s. Among other things they also assess where India stands with respect to the open economy Trilemma and how the Trilemma has evolved over the time period 1996Q1 to 2011Q3. Resonating the findings of the previous studies in this domain, their empirical results also establish that instead of adopting corner solutions, India has embraced an intermediate approach in managing the conflicting objectives of the Trilemma, balancing the three policy objectives as per the demands of the macroeconomic situation. They use the same measures for MI and KO as the previous studies but for ERS they use a different methodology introduced by Frankel and Wei (1994). In their method, the degree of influence of major global currencies on the Indian Rupee can be estimated using the following model:

$$\Delta\log \varepsilon_{\mathrm{INR},t}^{\mathrm{CHF}} = \alpha_0 + \beta_{\mathrm{US}}\Delta\log \varepsilon_{\mathrm{USD},t}^{\mathrm{CHF}} + \beta_{\mathrm{EUR}}\Delta\log \varepsilon_{\mathrm{EUR},t}^{\mathrm{CHF}} + \beta_{\mathrm{JAP}}\Delta\log \varepsilon_{\mathrm{JPY},t}^{\mathrm{CHF}} + \mu \tag{3}$$

Here, $\varepsilon_{i,t}^{\mathrm{CHF}}$ is the exchange rate of currency i against a numèraire currency, for which the authors use the Swiss franc and currency i can be the US Dollar, Japanese Yen and the Euro. If the Rupee is pegged to an individual or a basket of currencies the model will give a higher goodness of fit or R^2. Accordingly, the authors estimate this model for each quarter using daily data and take the adjusted R^2 as a measure of

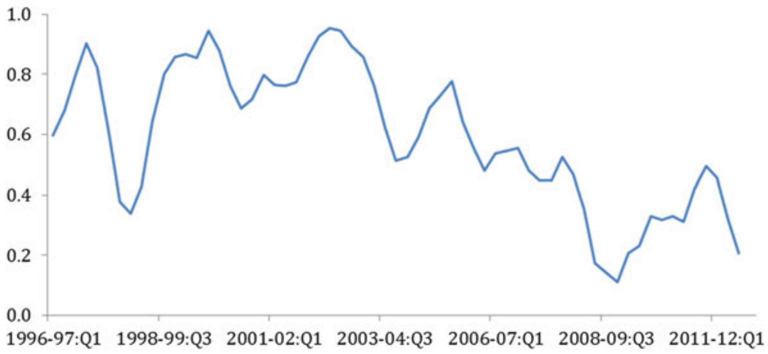
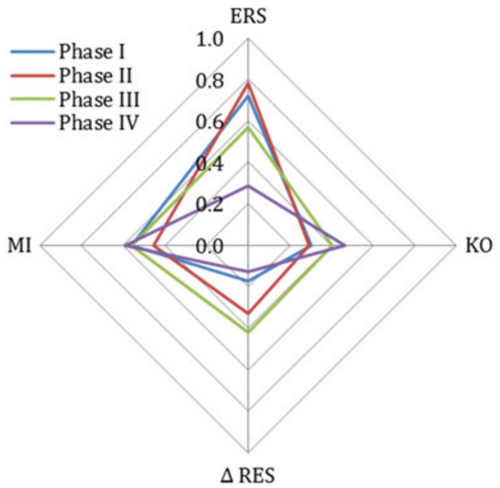


Fig. 3 Exchange rate stability index. *Source* Sen Gupta and Sengupta (2013)

ERS, with a higher R^2 indicating greater pegging to an individual or a basket of currencies. The evolution of this ERS series for the sample period considered by the authors is shown in Fig. 3.

The diamond chart from their study is shown in Fig. 4 and demonstrates the evolution of the Trilemma very clearly over the period from 1996Q1 to 2011Q3, split between four equal time windows. During phases II and III, i.e., from 2000Q1 to 20007Q4, RBI actively intervened in the foreign exchange market to prevent the Rupee from appreciating in the face of strong capital inflows. As mentioned earlier, while the RBI initially succeeded in sterilizing these interventions, from mid 2000s onwards rising fiscal cost of sterilization resulted in RBI only partially sterilizing its intervention, which in turn led to a reduction in monetary autonomy. In phase IV, i.e., from 2007Q4 to 2011Q3, there was discernible shift toward greater MI balanced with greater flexibility of the exchange rate, which has acted as a shock absorber in a period of volatile capital flows.

Fig. 4 Configuration of the trilemma objectives and international reserves. *Source* Sen Gupta and Sengupta (2013)



Thus, while the previous studies have concluded that MI was notably compromised for the greater part of the 2000s in favor of greater ERS and financial integration, this study finds that MI has been substantially restored at the expense of ERS toward the end of the 2000s. This is perhaps because their study extends till 2011Q3 and from mid 2010 onwards RBI took recourse to significant monetary tightening policies in order to deal with rising wholesale price inflation.

4 Concluding Remarks

The rise in the volatility of global capital flows has made macroeconomic management increasingly complex especially for emerging economies that are striving to achieve a balance between the diverse objectives of robust growth rate, sustainable current account deficit, competitive exchange rate, access to adequate external capital for financing investment, moderate inflation, and sufficient international reserves. This has also reignited the debate on suitable macroeconomic management measures. One of the main challenges that emerging economies face today as they integrate with global capital markets is managing the trade-offs presented by the well known Impossible Trinity or open economy Trilemma and in this, India is no exception. The results of the few select empirical studies that have attempted to quantify the various policy objectives under the Trilemma in India summarized here seem to be overall consistent with those found for a broader group of emerging economies. Among this group the policy combination of exchange rate stability and financial openness has been the most dominant over the past two decades.

India seems to have actively managed its exchange rate, building up a high level of international reserves by intervening heavily in the foreign exchange market and has also managed to retain some control over domestic monetary policy. In fact, active intervention in foreign exchange market coupled with maintaining a reasonable degree of control over international capital flows seem to have emerged as a potent combination of policy instruments in India as it strives to manage the macroeconomic Trilemma. These issues are likely to be highly relevant in the current context in India as policy makers dabble with inflation targeting to control domestic inflationary pressures, the Rupee exhibits renewed volatility in the face of foreign investment inflows and India strives to revive its growth rate.

References

Aizenman J (2010) The impossible trinity (aka the policy trilemma). In: Encyclopedia of financial globalization (May)

Aizenman J, Sengupta R (2013) The financial trilemma in China and a comparative analysis with India. Pacific Econ Rev 18(2):123–146

Aizenman J, Chinn M, Ito H (2008) Assessing the emerging global financial architecture: measuring the trilemma's configurations over time. NBER working paper no. 14533

Aizenman J, Chinn M, Ito H (2010a) The financial crisis, rethinking of the global financial architecture, and the trilemma. Asian Development Bank Institute Working Paper #213 (April)

Aizenman J, Chinn M, Ito H (2010b) The emerging global financial architecture: tracing and evaluating the new patterns of the trilemma's configurations. J Int Money Finance 29(4): 615–641

Aizenman J, Chinn M, Ito H (2011) Surfing the waves of globalization: asia and financial globalization in the context of the trilemma. J Japanese Int Econ 25(3):290–320

Chinn M, Ito H (2008) A new measure of financial openness. J Comp Policy Anal 10(3):307–320

Forbes KJ (2014) Capital flow volatility and contagion—a focus on Asia. In: Carrasco B, Gokarn S, Mukhopdhyay H (eds) Managing capital flows: issues in selected emerging market economies. Oxford University Press

Frankel JA, Wei SJ (1994) Yen bloc or dollar bloc? In: Ito T, Krueger AO (eds) Exchange rate policies of the east Asian economies. Macroeconomic linkages: savings, exchange rates, and capital flows. University of Chicago Press, pp 295–333

Hutchison M, Sengupta R, Singh N (2011) India's trilemma: financial liberalization, exchange rates and monetary policy. World Econ 35(1):3–18

Hutchison M, Kendall J, Pasricha G, Singh N (2012) Effectiveness of capital controls in India: evidence from the offshore NDF market. IMF Econ Rev 60(3):395–438

Lane P, Milessi-Ferreti GM (2007) The external wealth of nations mark II: revised and extended estimates of foreign assets and liabilities: 1970–2004. J Int Econ 73(2):223–250

Mohan R, Kapur M (2009) Managing the impossible trinity: volatile capital flows and Indian monetary policy. Stanford center for international development working paper no. 401

Obstfeld M, Shambaugh JC, Taylor AM (2005) The trilemma in history: tradeoffs among exchange rates, monetary policies, and capital mobility. Rev Econ Stat 87(3):423–438

Obstfeld M, Shambaugh JC, Taylor AM (2010) Financial stability, the trilemma, and international reserves. Am Econ J: Macroecon 2(2):57–94

Patnaik I, Shah A, Sethy A, Balasubramaniam V (2011) The exchange rate regime in asia: from crisis to crisis. Int Rev Econ Finance 20(3):32–43

Reddy YV (2008) Management of the capital account in India: some perspectives. Reserve Bank of India Bull February

Sen Gupta A, Sengupta R (2013) Managing capital flows in India. South Asia working paper series, Asian Development Bank, No. 17

Sen Gupta A, Sengupta R (2014) Negotiating the trilemma and reserve management in an era of volatile capital flows in India. In: Carrasco B, Gokarn S, Mukhopadhyay H (eds) Managing capital flows: issues in selected emerging market economies. Oxford University Press, pp 1–34

Sen Gupta A, Sengupta R (2015) Capital flows and capital account management in selected Asian economies. Forthcoming CIGI Working Paper

Shah A (2008) New issues in Indian macro policy. In: Ninan TN (eds) Business Standard India, Business Standard Books, Chap. 2, pp 26–54

Zeileis A, Shah A, Patnaik I (2010) Testing, monitoring, and dating structural changes in exchange rate regimes. Comput Stat Data Anal 54(6):1696–1706

Part VII
Issues Related to Financial Institutions

Guaranty Funds and Moral Hazard in the Insurance Industry: A Theoretical Perspective

C. William Sealey, John M. Gandar and Sumon C. Mazumdar

Abstract Third-party liability guarantees, such as the guarantees provided by state guaranty funds in the U.S. to property insurance companies, or federal deposit insurance provided to commercial banks in many countries, potentially create risk-taking incentives. Empirical evidence supports the risk-taking hypothesis in property liability insurance and banking. While a large body of literature has focused on the optimal design of regulations to address such moral hazard issues in banking, the potential risk-taking incentives that may arise in the property liability insurance context and their regulatory implications are much less understood. We develop a model that examines an insurance company's risk-taking incentives in the presence of a guaranty fund that protects policyholders from loss if the insurer fails, a scheme that is provided in many countries. We propose a set of workable regulatory mechanisms that may be considered to alleviate such potential moral hazard incentives.

1 Introduction

Given the central role that financial institutions play in the economy, the government (or a state agency) often provides such institutions with an explicit or implicit liability guarantee, i.e., promises to pay any remaining liability claims against the institution if the latter fails and is unable to fully pay such claims itself. Deposit insurance, designed to prevent bank runs, is perhaps the best known example of such a regulatory mechanism. Such insurance, provided by governments in many

C.W. Sealey · J.M. Gandar
University of North Carolina at Charlotte, Charlotte, USA

S.C. Mazumdar (✉)
Haas School of Business, University of California at Berkeley, Berkeley, USA
e-mail: sumon.mazumdar@navigant.com

S.C. Mazumdar
Navigant Economics, Oakland, USA

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_26

countries, has become even more wide spread after the 2007–2009 financial crisis.[1] However, it is well-understood that such a mechanism can increase a bank's incentive to "go for broke" (commonly referred to as moral hazard) unless the insurance mechanism is properly designed, for example through a risk-adjusted premium structure imposed on banks.[2] There is now a large body of literature that examines such moral hazard incentives in banking, and the design of banking regulation (e.g., risk-adjusted deposit insurance and capital requirements, or closure rules) to address such incentives.[3]

There is relatively less extant research about the moral hazard issues associated with the liability insurance that state guaranty funds provide insurance companies (i.e., the guarantee that the fund would protect policy holders if an insurance company in the state defaults on benefit payments or becomes insolvent) even though such insurance guarantee schemes which "provide last resort protection to policyholders and beneficiaries when insurers are unable to fulfil their contract commitments" are commonly offered across the world. According to an Oxera (2007) study, 13 of the 27 countries in the European Union had put in place insurance guarantee arrangements.[4] Such arrangements are in place in several other countries as well, such as the U.K., U.S., and Japan.[5]

Even though empirical evidence indicates that risk-taking by insurance companies is greater in the presence of such a insurance guarantee,[6] possible solutions to this risk-taking problem are much less understood. To date, policy proposals have been directed almost exclusively to the banking industry. However, the insurance business and its system of third-party guarantees are considerably different from its banking counterpart.

A bank's moral hazard incentive stems from its use of insured deposit debt to fund its assets. Such deposit insurance creates a possible moral hazard incentive if the deposit insurance agency (such as the United States' Federal Deposit Insurance

---

[1]See Demirgüç-Kunt et al. (2014) which provides a global database of deposit insurance arrangements as of 2013. The study also discusses the safety net provided in different countries to the financial system, which may include government protection of nondeposit liabilities and bank assets which increased during the recent crisis.

[2]Excessive risk-taking by banks resulting from deposit insurance-related moral hazard is considered among the primary factors that resulted in unprecedented deposit insurance losses during the 1980s and early 1990s [See Benston et al. (1986) for comprehensive discussion of this issue] Concerns about deposit insurance-related moral hazard continue today [See Demirgüç-Kunt et al. (2014) who note that "coverage of deposit insurance remains above precrisis levels, raising concerns about implicit coverage and moral hazard going forward."].

[3]See e.g., Chen et al. (2006), Giammarino et al. (1993), Yoon and Mazumdar (1996), Mazumdar and Yoon (1996) and Mazumdar (1997).

[4]"Insurance Guaranty Schemes In The EU: Comparative Analysis Of Existing Schemes, Analysis Of Problems And Evaluation of Options," Final report prepared for European Commission DG Internal market and Services, *Oxera Consulting Ltd.*, November 2007.

[5]See Kuras (2013).

[6]See, e.g., Lee et al. (1997) and Downs and Sommers (1999), and Ligon and Thistle (2008).

Corporation or FDIC) charges the bank an insurance premium that is too low given the bank's asset portfolio risk and capital position. Such mispriced deposit insurance can be viewed as an implicit subsidy that the bank and its shareholders get from the government and ultimately, taxpayers. The bank's shareholders then have a moral hazard incentive to increase this subsidy's value by increasing the bank's asset portfolio risk.[7]

In contrast, the source of an insurance company's moral hazard incentive is not its use of debt financing. Instead, even a wholly equity financed insurer may face a moral hazard incentive given an insurance guarantee scheme, as we describe in greater detail later. Among other differences between deposit insurance and an insurance guarantee scheme, the insurance guaranty scheme may not charge the insurer a premium. Instead, all solvent insurers in the guaranty fund's jurisdiction may be required to collectively cover shortfalls of all insolvent insurers in the jurisdiction, with each solvent insurer contributing in proportion to its size, measured by its relative size of total premiums collected by insurers in the jurisdiction. Therefore, mechanisms proposed in the banking context to alleviate bank moral hazard (such as capital requirements, monitoring rules and actuarially fair deposit insurance pricing) are not directly applicable to the insurance industry.

To focus on a representative property insurance company's (or insurer's) risk-taking incentives in the presence of a insurance guarantee scheme, we develop a model that examines two particular decisions the insurer must make, *namely,* (i) its choice of assets in which to invest the premiums it collects (i.e., its *asset quality* choice); and (ii) the losses it chooses to insure (i.e., its *underwriting quality* choice). Our model examines the insurer's relationship to regulators, the guaranty fund, and policyholders in an asymmetric information setting and highlights the factors that affect the insurer's risk-taking choices. We next discuss a set of workable mechanisms that could mitigate such moral hazard problems associated with property insurance.

Before discussing our model and its implications a comment is in order. In practice, the factors that affect a particular insurance company's risk-taking activity may be more complex than those addressed in our theoretical model. Nevertheless, we hope that our research may offer insights that guide future research and policy proposals that address potential risk-shifting in the insurance industry. In that connection it is worth noting that although our discussion is motivated by the moral hazard issues that property insurers face, our analysis may also shed light on the

[7]The intuition behind this result is straightforward. A riskier bank asset portfolio means that the bank's future asset values are potentially higher in certain ("favorable") states and worse in other "unfavorable" states, when the bank is forced to default on its deposit debt. However, given limited liability the bank's shareholders losses are limited to zero. Therefore, bank risk-taking increases the shareholders' payoffs in favorable states without any incremental loss in unfavorable states. Instead the deposit insurer's burden (to pay off the bank's depositors after the bank defaults) increases. But if deposit insurance is not actuarially fair the insurer is not compensated by the bank for bearing this additional risk. Hence, holding other factors constant, the greater the bank risk the greater is the deposit insurance mispricing (or size of the deposit insurance subsidy) and gains to the bank's shareholders.

design of insurance guarantee schemes related to other types of insurance companies that share certain characteristics of the insurer and insurance guarantee scheme we have modeled.[8]

2 The Model

We develop a model that examines the insurance company's interaction with policyholders and the guaranty fund under conditions of moral hazard. The model's results can be classified into three general categories

1. As a benchmark, we develop a model of insurer-policyholder interaction from the point of view of full information, no regulation, and no third-party guarantees. The insurer's optimal risk level, where the insurer acts in the interest of equity holders, is determined in this unregulated environment, and then compared to the socially optimal level of risk.
2. Next, we introduce a guaranty fund and examine the risk levels the insurer chooses when its asset and underwriting quality are private information and cannot be freely observed. Insurers are free to take advantage of their private information. We demonstrate that the insurer has an incentive to increase its risks (i.e., lower its asset and underwriting qualities) in the presence of a guaranty fund.
3. Finally, we examine regulatory mechanisms involving state-contingent payments that the insurer must make to the guaranty fund that may alleviate the moral hazard problem. We demonstrate that such mechanisms, if suitably calibrated, could lead an insurer to choose asset and underwriting qualities that are superior from the social planner's point of view, even in a world with informational asymmetry.

2.1 The Setup for the Insurer

Consider a simple one period model of a property liability insurance company (or insurer). At the beginning of the period (which we refer to as date $t = 0$), the representative insurer obtains paid-in equity capital, E, and underwrites liability insurance from which it collects premiums. Equity is assumed to be restricted to a specified α percent of premiums collected. For simplicity, it is assumed that the insurer invests all available funds (equity plus premiums) in risky assets, A. The

[8]Such key characteristics include: solvent insurance companies collectively bearing the loss of other insolvent insurers and insurance companies not being required to pay a risk-adjusted premium for participating in the insurance guarantee scheme. We thank the referee for pointing out the potentially broader applicability of our theoretical framework.

underwriting activities generate liabilities to the insurer which may, but need not, be constrained by capital requirements or other restrictions set by a regulator. The insurer's managers are assumed to make decisions in the interest of shareholders. All parties in the model are risk neutral, and, for simplicity, the risk-free rate of return is assumed to be zero, i.e., the discount factor is equal to one.

At $t = 0$, the insurer faces a continuum of asset and underwriting options that differ only in terms of quality. Let $q_A \in (0, 1)$ denote the quality of the insurer's asset portfolio and let $q_L \in (0, 1)$ denote the quality of its underwriting liabilities. The insurer's management chooses q_A and q_L, *ex ante* (at $t = 0$). The (gross) return on the insurer's asset portfolio (principal plus return) and any losses from underwriting activities are realized at the end of the period (i.e., at date $t = 1$). Such *ex post* outcomes depend, in the first instance, on the insurer's *ex ante* quality decisions q_A and q_L.

Outcomes for both asset returns and underwriting losses are assumed to be binomial. The return on assets will be favorable (relatively high) with probability q_A, or unfavorable (low) with probability $(1 - q_A)$. Thus, higher levels of q_A indicate higher asset quality. If the outcome is favorable, the insurer's return per dollar of assets is denoted by $R(q_A) + \tilde{\varepsilon}_A$, where $\tilde{\varepsilon}_A$ is unique, asset-specific noise and $E(\tilde{\varepsilon}_A) = 0$. If the outcome is unfavorable, the return on the asset portfolio is a fixed percentage of assets, R_2. The expected return on assets in the favorable state is, of course, greater than the return in the unfavorable state, i.e., $[R(q_A) + E(\tilde{\varepsilon}_A)] > R_2 \ \forall\, q_A \in (0, 1)$. Henceforth, to simplify notation, the expectation of $R(q_A)$ over $\tilde{\varepsilon}_A$ will be denoted by $\overline{R}(q_A)$.

For underwriting losses, if the outcome is favorable, losses are relatively small. The probability of the favorable state is q_L. If the outcome for losses is unfavorable, losses are high, with probability $(1 - q_L)$. Thus, higher levels of q_L indicate higher underwriting quality. If a favorable outcome occurs then the insurer incurs a loss rate (per dollar of assets) $L(q_L) + \tilde{\varepsilon}_L$, where $\tilde{\varepsilon}_L$ is unique, underwriting specific noise and $E(\tilde{\varepsilon}_L) = 0$.[9] L_2 denotes the (fixed) underwriting loss rate per dollar of assets in the unfavorable state. Underwriting losses in the favorable state are always expected to be less than losses in the unfavorable state, i.e., $L_2 > L(q_L) + E(\tilde{\varepsilon}_L) \ \forall\, q_L \in (0, 1)$, where $L(q_L) + E(\tilde{\varepsilon}_L)$ is hereafter denoted as $\overline{L}(q_L)$. Moreover, expected losses conditional on the favorable state are assumed not to be sufficiently great to bankrupt the insurer, irrespective of asset returns, i.e., $\overline{L}(q_L) \le R_2$, $\forall\, q_L \in (0, 1)$. Finally, $L_2 > R_2$, which means that if the insurer's assets earn unfavorable (low) returns and the insurer simultaneously experiences an unfavorable large underwriting loss, the insurer will become insolvent, i.e., in that state of the world the insurer's loss will exceed the value of its assets at the end of the period (at date $t = 1$).

The asset return function, $R(q_A)$, is such that $R'(q_A) < 0$. That is, a higher asset quality level (a higher q_A) increases the likelihood of a favorable return on investments in assets, but also reduces the magnitude of the return in the favorable state. In

[9]The random noise terms $\tilde{\varepsilon}_L$ and $\tilde{\varepsilon}_A$ are assumed to be uncorrelated.

addition, the expected return in the favorable state, $q_A R(q_A)$, increases with better quality, but at a decreasing rate, (i.e., $q_A R(q_A)$ is increasing and concave in q_A).

The loss function, $L(q_L)$, is assumed to be such that $L'(q_L) < 0$,[10] while $q_L L(q_L)$ is decreasing and convex in q_L. That is, higher underwriting quality lowers expected losses in the favorable loss state, but at a decreasing rate.

The insurer's choice of q_A and q_L are private information, and cannot be observed, even *ex post*. The presence of noise in asset returns and underwriting losses, $\tilde{\varepsilon}_A$ and $\tilde{\varepsilon}_L$, ensures that neither q_A nor q_L can be inferred from the *ex post* realization of returns. This gives rise to the usual moral hazard problem in that the insurer might have an incentive to choose asset and underwriting quality levels that may be suboptimal from a social or regulatory point of view. That is, an equity value maximizing insurer may choose to invest in lower quality assets that offer greater upside return albeit with greater risk, and/or underwrite lower quality risks with a higher probability of experiencing a larger loss, if it results in higher expected returns to its shareholders who enjoy the benefit of limited liability. Such risk-taking incentives may be exacerbated when the insurer also receives assurance from the guaranty fund that if it fails, the fund would cover its policyholders' losses. Nevertheless, it may be possible for the insurer's quality decisions, q_A and q_L, to be controlled through regulatory design as we discuss in later sections of this chapter.

2.2 *The Insurer's State-Contingent Payoffs*

Given the binomial outcomes of asset returns and underwriting losses, there are four possible, mutually exclusive, states that may be realized at $t = 1$. The random return per dollar of assets for the insurer at $t = 1$, is one of the following:

State 1: The outcome of the asset investment and the underwriting loss are both favorable. The insurer's net payoff (net of its underwriting losses) is $[R(q_A) + \tilde{\varepsilon}_A] - [L(q_L) + \tilde{\varepsilon}_L]$, which occurs with probability $q_A q_L$. In this state, $\overline{R}(q_A) > \overline{L}(q_L)$. The insurer is expected to be solvent.

State 2: The outcome of the asset investment is favorable but the underwriting loss is unfavorable. The insurer's net payoff is $[R(q_A) + \tilde{\varepsilon}_A] - L_2$, which occurs with probability $q_A(1 - q_L)$. In this state, $\overline{R}(q_A) > L_2$. The insurer is expected to be solvent.

State 3: The outcome of the asset investment is unfavorable but the underwriting loss is favorable. The insurer's net payoff is $R_2 - [L(q_L) + \tilde{\varepsilon}_L]$, which occurs with probability $(1 - q_A)q_L$. In this state, $R_2 > \overline{L}(q_L)$. The insurer is expected to be solvent.

[10] In addition, it seems reasonable to assume that $L''(q_L) > 0$, i.e., as underwriting quality decreases, losses increase at a decreasing rate.

State 4: The outcomes of the asset investment as well as the underwriting loss are both unfavorable. The insurer's net payoff is $R_2 - L_2$, which occurs with probability $(1 - q_A)(1 - q_L)$. In this state, as $(R_2 - L_2) < 0$, the insurer is insolvent. Its equity holders receive nothing. The insurer's assets are divided among its various policyholder claimants according to prespecified priority rules, which are assumed for the time being to be pro-rated, based on premiums paid by policyholders. A shortfall in the payments to policyholders may be paid by a guaranty fund or reinsurer where such alternatives exist.

3 Optimal Quality Decisions from the Social Planner's and the Insurer's Perspectives

In this section, we examine two benchmark cases where the quality decisions of the insurer are observable. Constraints on insurer behavior (such as regulatory constraints) are assumed to be absent and, in any case, are unnecessary. First, we determine the socially optimal, first-best levels for the insurer's asset quality and underwriting qualities, q_A and q_L, i.e., the asset quality and underwriting quality levels which maximize social surplus. Such a socially optimal allocation is useful for the purpose of comparing the benefits of the regulatory mechanism that is considered later. Second, we determine the optimal asset and underwriting quality decisions from the insurer's perspective where the insurer is assumed to act to maximize the value of its equity.

3.1 Underwriting Premiums

Premiums are collected at $t = 0$. The insurer's available funds for investments in assets are the sum of premiums collected and paid-in equity capital. Given the assumption of risk neutrality, the maximum aggregate premium that an insurer can charge is the present value of the expected recoverable losses by policyholders.[11] In turn, expected recoverable losses are the policyholders' total expected losses less that part of losses that are unrecoverable in the event that the insurer is insolvent at $t = 1$ which occurs in *State 4*, as we discussed earlier. $P(q_A, q_L)$ denotes total premiums collected by the insurer, and is a function of both q_A and q_L, because policyholders' recoverable losses depend on the insurer's underwriting quality as well as the quality of its invested assets. The size of the insurer's asset portfolio at

[11]Obviously, even if policy holders are risk averse, as long as insurers are risk neutral they assume all risk and premiums are competitively priced at the level of expected loss plus operating costs. For simplicity, operating costs are assumed to be zero.

$t = 0$ equals its total premiums plus its paid-in capital which is assumed to be α times its total premiums. As total premiums are a function of the insurer's asset and underwriting quality choices, it follows that the size of the insurer's asset portfolio, $A(q_A, q_L)$, depends on its quality decisions, i.e., $A(q_A, q_L) \equiv P(q_A, q_L)(1+\alpha)$. For the assumed discount factor of one, and in the absence of a guaranty fund or reinsurer, total premiums can be expressed as follows:

$$P(q_A, q_L) = [(1-q_L)L_2 + q_L L(q_L)] - [(1-q_L)(1-q_A)(L_2 - R_2)], \qquad (1)$$

where the first term in brackets on the RHS is the total expected losses of policyholders, and the second term in brackets is the expected unrecoverable losses by policyholders as a result of insurer insolvency. Other things equal, the greater the expected unrecoverable losses, the smaller the premiums that policyholders are willing to pay. On the other hand, the greater policyholders' expected total losses, the greater is the premium they are willing to pay.

As Eq. (1) suggests, since asset quality impacts only expected unrecoverable losses to policyholders, it is clear that the premiums collected by the insurer are increasing in asset quality, since expected unrecoverable losses are declining at a linear rate, i.e., $\partial P(q_A, q_L)/\partial q_A > 0$ and $\partial^2 P(q_A, q_L)/\partial q_A^2 = 0$.

The first derivative of (1) with respect to q_L is

$$\frac{\partial P(q_A, q_L)}{\partial q_L} = \left(\frac{\partial q_L L(q_L)}{\partial q_L}\right) - q_A L_2 - (1-q_A)R_2 < 0.$$

The second derivative is $\frac{\partial^2 P(q_A, q_L)}{\partial q_L^2} = \left(\frac{\partial^2 q_L L(q_L)}{\partial q_L^2}\right) > 0$. Thus, underwriting premiums are decreasing and convex in q_L. These properties are intuitively appealing as expected losses should decline as underwriting quality increases, but losses decrease at a decreasing rate.

3.2 The Social Planner's Problem

The social planner seeks to maximize expected social surplus.[12] Expected social surplus, denoted by, $\pi^{SP}(q_A, q_L)$, is given by

[12]Here, we assume that the social planner considers social surplus to be composed only of the expected monetary returns to all insurer claimholders. In the case of insurance companies, one might argue that society in general benefits from high-quality decisions by insurers through the positive externalities of financial system stability. Although difficult to quantify, if such externalities were included, the social surplus generated by high quality would increase and the first-best quality chosen by the social planner over that implied by the first-order conditions in Eqs. (3) and (4).

$$\pi^{SP}(q_A, q_L) = \{q_A q_L[R(q_A) - L(q_L)] + q_A(1 - q_L)[R(q_A) - L_2] + (1 - q_A)q_L[R_2 - L(q_L)] \\ - (1 - q_A)(1 - q_L)[L_2 - R_2]\}A(q_A, q_L). \tag{2}$$

Defining $\xi^{SP}(q_A^{FB}, q_L)$ as the social surplus *per dollar* of insurer assets, the RHS of the above expression can be rewritten as $\xi^{SP}(q_A^{FB}, q_L)\, A(q_A, q_L)$.

The social planner's problem is to choose q_A and q_L to maximize (2). The social planner's optimal choices of quality, denoted by q_A^{FB} and q_L^{FB}, are given by the first-order conditions, which after rearranging, are as follows

$$\frac{\partial q_A^{FB} R(q_A^{FB})}{\partial q_A^{FB}} = R_2 - \frac{[\xi^{SP}(q_A^{FB}, q_L)]P'_{q_A}(q_A^{FB}, q_L)}{P(q_A^{FB}, q_L)}, \tag{3}$$

and

$$\frac{\partial q_L^{FB} L(q_L^{FB})}{\partial q_L^{FB}} = L_2 + \frac{[\xi^{SP}(q_A, q_L^{FB})]P'_{q_L}(q_A, q_L^{FB})}{P(q_A, q_L^{FB})}. \tag{4}$$

It can be shown that the second derivatives with respect to q_A and q_L are negative.

3.3 *The Insurer's Problem*

Now, consider the optimal asset and underwriting quality decisions the insurer makes in order to maximize the value of its equity in the presence of limited liability. Let $\pi^E(q_A, q_L)$ denote the value of equity. The insurer, acting in the interest of equity holders, maximizes the following:

$$\pi^E(q_A, q_L) = \{q_A q_L[R(q_A) - L(q_L)] + q_A(1 - q_L)[R(q_A) - L_2] \\ + q_L(1 - q_A)[R_2 - L(q_L)]\}A(q_A, q_L). \tag{5}$$

Defining $\xi^E(q_A^{SB}, q_L)$ as the insurer's equity value *per dollar* of assets, the RHS of the above expression can be re-written as $\xi^E(q_A^{SB}, q_L)\, A(q_A, q_L)$.

The insurer's choice of quality, which are second best quality decisions are denoted as q_A^{SB} and q_L^{SB}, respectively. These levels are given by the first-order conditions

$$\frac{\partial q_A^{SB} R(q_A^{SB})}{\partial q_A^{SB}} = (1 - q_L)L_2 + q_L R_2 - \frac{[\xi^E(q_A^{SB}, q_L)]P'_{q_A}(q_A^{SB}, q_L)}{P(q_A^{SB}, q_L)}, \tag{6}$$

and

$$\frac{\partial q_L^{SB} L(q_L^{SB})}{\partial q_L^{SB}} = (1 - q_A)R_2 + q_A L_2 + \frac{[\xi^E(q_A,\, q_L^{SB})]P'_{q_L}(q_A,\, q_L^{SB})}{P(q_A,\, q_L^{SB})}. \tag{7}$$

It can also be shown that the second derivatives with respect to q_A and q_L are negative.

3.4 Optimum Quality Comparisons for the Social Planner and Equity Holders

Based on the above first-order conditions, it is possible to compare the quality decisions made by the social planner and the insurer acting to maximize its own equity value. Such a comparison yields the following results.

Proposition 1 *When the quality decisions of the insurer are observable, then* $q_A^{FB} > q_A^{SB}$ *and* $q_L^{FB} > q_L^{SB}$ *for* $\forall\, q_L \in (0, 1)$ *and* $\forall\, q_A \in (0, 1)$, *respectively.*

Proof See appendix.

In words, the insurer's choice for both asset and underwriting quality in the absence of moral hazard is unambiguously lower than the social optimum (or "second-best"). This result reflects the expected impact of limited liability on decision making. It is important to note, however, that even though the insurer prefers lower asset and underwriting quality (compared to the social optimal), there is no shifting of risk under this full information scenario because the premiums that policyholders are willing to pay the insurer reflects the insurer's quality choices.

4 Insurer's Quality Choices in the Presence of a Guaranty Fund with a Pro-Rata Loss Sharing Scheme

In this section, we first examine the quality decisions of insurers when the guaranty fund acts in a passive manner, i.e., the fund does not monitor or regulate the insurer. The only restriction the insurer faces is that its equity capital is restricted to a fixed portion of premiums, as noted earlier.

Under these conditions, the insurer faces a somewhat different environment from the one modeled above. First, premiums are as follows:

$$P^{GF}(q_L) = (1 - q_L)L_2 + q_L L(q_L), \tag{8}$$

which is equal to the policyholders' expected losses. Comparing this equation to (1), which denoted the total expected premiums in the absence of a guaranty fund, indicates that the only difference is that the second term on the RHS of Eq. (1) is missing in the above expression. That is, given the presence of a guaranty fund that covers any shortfall to the policyholder if the insurer's assets were insufficient to cover all claims, policyholders' expected *unrecoverable* losses as a result of insurer insolvency are now zero. As a result, premiums are only a function of underwriting quality, q_L. The insurer's asset quality, q_A is immaterial because there are no unrecoverable losses.

In the U.S., if an insurer's assets are inadequate to meet its policyholders' claims, the state guaranty fund under whose jurisdiction this insolvent insurer belongs covers the shortfall so that the policyholders do not bear any unrecovered loss. Shortfalls (or loss) faced by insolvent insurers are paid by the remaining solvent insurers within the jurisdiction, with each solvent insurer's share of the total insurance losses being proportional to that insurer's share of the total premiums collected within the state. As the guaranty fund does not charge any risk-based insurance premium for providing insurance companies within its jurisdiction a liability guarantee, an equity value maximizing insurance company may have an incentive for risk-taking, absent other constraints, as we analyze below.

The payoff to an individual insurer's equity holders in the presence of a guaranty fund is given by

$$\begin{aligned}\pi^E_{GF}(q_A, q_L) = \{&q_A q_L[R(q_A) - L(q_L)] + q_A(1-q_L)[R(q_A) - L_2] \\ &+ q_L(1-q_A)[R_2 - L(q_L)]\}A^{GF}(q_A, q_L) - \gamma TIL.\end{aligned} \tag{9}$$

where $A^{GF}(q_L) \equiv P^{GF}(q_L)(1+\alpha)$. The last term on the RHS of (9) indicates the insurer's total contribution to the state guaranty fund. It is a product of γ (the insurer's pro-rata share of the total premiums collected in the state) and *TIL* (total insurance losses to all insurance policyholders within the guaranty fund's jurisdiction due to insurer failures). In the present setup, it is immaterial whether the insurer is large or small relative to the insurance market.

With the introduction of the guaranty fund, the quality pair that achieves a socially optimal allocation remains (q_A^{SP}, q_L^{SP}). The equity value-maximizing insurer chooses the quality pair (q_A^{GF}, q_L^{GF}) in the presence of the guaranty fund. Such choices are given by the two first-order conditions

$$\frac{\partial q_A^{GF} R(q_A^{GF})}{\partial q_A^{GF}} = (1-q_L)L_2 + q_L R_2, \tag{10}$$

and

$$\frac{\partial q_L^{GF} L(q_L^{GF})}{\partial q_L^{GF}} = (1-q_A)R_2 + q_A L_2 + \frac{[\xi^E(q_A, q_L^{GF})]P^{GF'}(q_L^{GF})}{P^{GF}(q_L^{GF})}. \tag{11}$$

It can also be shown that the second derivatives with respect to q_A and q_L are negative.

As noted above, the insurance company is not charged a risk-based premium for obtaining a liability guarantee from the guaranty fund. The presence of such a guarantee affects the insurer's decisions about q_A and q_L. To assess the impact of such a liability guarantee structure on the insurance company's optimal (a) asset quality decision one must compare the first-order conditions in Eqs. (6) and (10); and (b) underwriting quality decision one must compare the first-order conditions in Eqs. (7) and (11). Such comparisons yield the following proposition.

Proposition 2 *When the quality decisions of the insurer are private information, and in the presence of a guaranty fund, then* $q_A^{SB} > q_A^{GF}$ *and* $q_L^{SB} > q_L^{GF}$ *are unambiguously true* $\forall\, q_L \in (0,1)$ *and* $\forall\, q_A \in (0,1)$, *respectively.*

Proof **See appendix.**

In short, under a guaranty fund scheme that prices according to a pro-rata loss rule, the insurer has an incentive to increase both asset risk and underwriting risk, compared to the quality decisions it would have made as an equity value maximizing firm in the absence of a guaranty fund.

5 An Alternative Mechanism for Mitigating the Moral Hazard Problem

In this section, we consider an alternative contracting mechanism, which takes the form of an option issued by the insurer to the guaranty fund at $t = 0$ (in addition to its pro-rata loss share) for the liability guarantee it obtains from the fund. This option entitles the guaranty fund to receive certain additional (state-contingent) payments *ex post* from the insurer in the solvent states (1, 2, and 3). We show that this alternative regulatory design can significantly mitigate the moral hazard problem, provided the option is optimally designed.[13]

5.1 Optimal Asset Quality When Underwriting Quality Is Given and Guaranty Fund Holds an Option on Insurer's Future Cash Flow

Let $\bar{q}_L$ denote a fixed value of underwriting quality and C_A^{GFO} denote the option payoff to the guaranty fund at $t = 1$, which can be written as

$$C_A^{\mathrm{GFO}} = \max\{\lambda_A[q_A R(q_A)(1+\alpha)P(q_A,\bar{q}_L) - (1-\eta_A)\Phi_A] - \eta_A\Phi_A,\, 0\}, \tag{12}$$

[13]The solution discussed in this section is in the spirit of John et al. (1991), Green (1984).

where $0 < \eta_A \leq 1$, and $0 \leq \lambda_A \leq 1$ are constants set by the guaranty fund[14] and $\Phi_A \equiv (1-q_A)(1-\bar{q}_L)R_2 - (1-\bar{q}_L)L_2 - q_A\bar{q}_L L(\bar{q}_L)$. In effect, the guaranty fund receives an option with an exercise price that is a portion, η_A, of the value of Φ_A. This option might be viewed as a payment to the guaranty fund for the guarantee it provides. There are optimal values for η_A and λ_A that not only mitigates the moral hazard problem but may also achieve a first-best solution.

Given the above features, the guaranty fund optimally exercises the asset quality option if and only if $\lambda_A[q_A R(q_A)(1+\alpha)P(q_A,\bar{q}_L) - (1-\eta_A)\Phi_A] - \eta_A\Phi_A \geq 0$.

Now, define an indicator function

$$1_A^{\mathrm{GFO}} \equiv \begin{cases} 1 & \text{if} \quad \lambda_A[q_A R(q_A)(1+\alpha)P(q_A,\bar{q}_L) - (1-\eta_A)\Phi_A] - \eta_A\Phi_A \geq 0 \\ 0 & \text{otherwise;} \end{cases}.$$

whose value depends on whether or not the option is in-the-money. The expected payoff to the insurer's equity holders for a given value of $\bar{q}_L$ can be written as

$$\begin{aligned} \pi^E(q_A,) = {} & q_A R(q_A)(1+\alpha)P(q_A,\bar{q}_L) + (1-q_A)(1-\bar{q}_L)R_2 \\ & - (1-\bar{q}_L)L_2 - q_A\bar{q}_L L(\bar{q}_L) - C_A^{\mathrm{GFO}} \times 1_A^{\mathrm{GFO}} - \gamma TIL. \end{aligned} \tag{13}$$

The first-order condition for the insurer's optimal asset quality, q_A^*, that maximizes shareholder value, given that underwriting quality is fixed, is as follows:

$$\begin{aligned} & \left[\frac{\partial q_A R(q_A)]}{\partial q_A}(1+\alpha)P(q_A,\bar{q}_L) - \frac{\partial \Phi_A}{\partial q_A}\right] \\ & - \left\{\lambda_A\left[\frac{\partial q_A R(q_A)]}{\partial q_A}(1+\alpha)P(q_A,\bar{q}_L) - (1-\eta_A)\frac{\partial \Phi_A}{\partial q_A}\right] - \eta_A\frac{\partial \Phi_A}{\partial q_A}\right\} \times 1_A^{\mathrm{GFO}} = 0. \end{aligned} \tag{14}$$

The condition in (14) leads to the following result:

Proposition 3 *Let* $\Omega_A \equiv R_2 - q_A^*(1+\alpha)R(q_A^*)P'_{q_A}(q_A^*,\bar{q}_L)$. *The optimal option contract for the guaranty fund requires values for* λ_A^* *and* η_A^* *as follows*

1. $\lambda_A^* \geq \left[\frac{\eta_A(\partial\Phi_A/\partial q_A^*)}{\frac{\partial q_A^* R(q_A^*)]}{\partial q_A^*}(1+\alpha)P(q_A^*,\bar{q}_L) - (1-\eta_A)\frac{\partial\Phi_A}{\partial q_A^*}}\right]$; and
2. $\eta_A^* = \left[\frac{(\partial\Phi_A/\partial q_A^*) - \Omega_A}{(\partial\Phi C_A/\partial q_A^*)}\right]$.

[14] As will be evident later in this section, neither $0 = \eta = 1$ and $\lambda = 0$, nor $0 = \eta = 1$ and $0 = \lambda = 1$ are possible optimal solutions for the model. In the former case, the guaranty fund is redundant, and in the latter the option is always worthless.

The above values for λ_A^* and η_A^* in turn induce the insurer to choose optimal asset quality, q_A^*, such that

$$\left[\frac{\partial q_A^* R(q_A^*)]}{\partial q_A^*}(1+\alpha)P(q_A^*, \bar{q}_L)\right] - (1-\eta_A^*)\frac{\partial \Phi_A}{\partial q_A^*} = 0.$$

Furthermore, the insurer's optimal asset quality, q_A^*, achieves the first-best level of quality, i.e., $q_A^* = q_A^{FB}$.[15]

5.2 *Optimal Underwriting Quality When Asset Quality Is Given and Guaranty Fund Holds an Option on Insurer's Future Cash Flow*

Let C_L^{GFO} denote the expected option payoff to the guaranty fund at $t = 1$, given q_A is fixed. This terminal payoff can be written as

$$C_L^{\text{GFO}} = \max\{\lambda_L[q_L L(q_L) - (1-\eta_L)\Phi_L] - \eta_L \Phi_L,\ 0\}, \tag{15}$$

where $0 < \eta_L \le 1$, and $0 \le \lambda_L \le 1$ are constants that are set by the guaranty fund,[16] and $\Phi_L \equiv \bar{q}_A R(\bar{q}_A)(1+\alpha)P(\bar{q}_A, q_L) + (1-\bar{q}_A)(1-q_L)R_2 - (1-q_L)L_2$.

The guaranty fund optimally exercises the underwriting quality option if and only if $\lambda_L[-\bar{q}_A q_L L(q_L) - (1-\eta_L)\Phi_L] - \eta_L \Phi_L \ge 0$. Now, define the indicator function

$$1_L^{\text{GFO}} \equiv \begin{cases} 1 & \text{if } \lambda_L[q_L L(q_L) - (1-\eta_L)\Phi_L] - \eta_L \Phi_L \ge 0 \\ 0 & \text{otherwise.} \end{cases}.$$

The expected payoff to stockholders for a given value of $\bar{q}_A$ can be written as

$$\begin{aligned} \pi^E(\bar{q}_A, q_L) = {} & \bar{q}_A R(\bar{q}_A)(1+\alpha)P(\bar{q}_A, q_L) + (1-\bar{q}_A)q_L R_2 \\ & - (1-q_L)\bar{q}_A L_2 - q_L L(q_L) - C_L^{\text{GFO}} \times 1_L^{\text{GFO}} - \gamma TIL. \end{aligned} \tag{16}$$

[15]The proof of this Proposition can easily be verified by substituting λ_A^* and η_A^* into (18) and simplifying. The resulting first-order condition is the same as (3), yielding the first-best asset quality.

[16]Analogous to the former case, $0 = \eta_L = 1$ and $0 = \lambda_L = 1$ are not possible optimal equilibrium solutions for the model.

The first-order condition for equity maximization is

$$\left[\frac{\partial q_L L(q_L)]}{\partial q_L} - \frac{\partial \Phi_{q_L}}{\partial q_L}\right] - \left\{\lambda_L\left[\frac{\partial q_L L(q_L)]}{\partial q_L} - (1-\eta_{q_L})\frac{\partial \Phi_{q_L}}{\partial q_L}\right] - \eta_{q_L}\frac{\partial \Phi_{q_L}}{\partial q_L}\right\} \times 1_L^{\text{GFO}} = 0. \tag{17}$$

Proposition 4 *Let* $\Omega_L \equiv \frac{L_2-(1-\bar{q}_A)R_2-\bar{q}_A(1+\alpha)R(\bar{q}_A)P'_{q_L^*}(\bar{q}_A,q_L^*)}{\bar{q}_A}$. *The optimal option contract for the guaranty fund requires values for* λ_L^* *and* η_L^* *as follows:*

1. $\lambda_L^* \geq \left[\frac{\eta_L(\partial\Phi_L/\partial q_L^*)}{\frac{\partial[q_L L(q_L^*)]}{\partial q_L^*}-(1-\eta_L)\frac{\partial\Phi_{q_L}}{\partial q_L^*}}\right]$; and
2. $\eta_L^* = \left[\frac{(\partial\Phi_L/\partial q_L^*)-\Omega_L}{(\partial\Phi_L/\partial q_L^*)}\right]$.

The above values for λ_L^* *and* η_L^* *in turn induce the insurer to choose optimal underwriting quality,* q_L^*, *such that*

$$\left[\frac{\partial q_L^* R(q_L^*)]}{\partial q_L^*}\right] - (1-\eta_{q_L}^*)\frac{\partial \Phi_L}{\partial q_L^*} = 0.$$

Furthermore, the insurer's optimal underwriting quality, q_L^*, achieves the first-best level of quality, i.e., $q_L^* = q_L^{FB}$.[17]

5.3 *Discussion of the Contingent Regulatory Mechanism*

An examination of the contingent payoffs that the insurer's shareholders and the guaranty fund receive at the end of the period reveals an important feature of the regulatory design. In particular, if the contract is designed optimally (as shown in Propositions 3 and 4) then in equilibrium the insurer chooses first-best quality, and the guaranty fund's option is not expected to be exercised because it is expected to be exactly at-the-money at $t = 1$.[18] It is also interesting to note that the contract is, in principle, similar to an *ex post*, risk-adjusted guaranty fund pricing scheme.

The overall intuition behind our results is as follows. The introduction of a guaranty fund (coupled with a pro-rata loss rule) induces the insurer to increase its asset and underwriting risks as the net payoffs to its shareholders at the end of the

[17]The proof of Proposition 4 is analogous to that of Proposition 3.

[18]Because of asset-specific risk, the guaranty fund options could be in-the-money in spite of the bank choosing first-best quality. On average, however, the institution breaks even if rebates are given when over-payments are made. See Nagarajan and Sealey (1995, 1998).

period are then higher in some states of the world, while not lower in other states. One way such excessive risk-taking may be prevented is by forcing the insurer to share the upside it generates by its excessive risk-taking with the guaranty fund. In effect, the guaranty fund holds a option to claw back the insurer's excessive gains. If such a claw back feature is optimally designed *ex ante* then it is never necessary to exercise the option *ex post* because its mere presence, which constitutes a contingent liability from the insurer's perspective, will induce the insurer to behave appropriately at the beginning of the period when it must choose its asset and underwriting risks.

6 Conclusion

Although explicit or implicit liability guarantees are often provided to financial institutions by the government or a state agency to ensure the stability of the financial system, such guarantees may lead the financial institution to undertake excessive risk relative to the socially optimal level. Such a moral hazard incentive arises because the potential loss associated with such risk is borne by the government while the financial institution's owners (its equity holders) receive the additional potential upside gain associated with such risk-taking. A large and expanding body of academic research continues to study such risk-taking incentives associated with deposit insurance provided to banks. However, there has been relatively little research on the moral hazard incentives, albeit of a different nature, that may also arise in the property liability insurance sector where state guaranty funds cover any loss to the insurance company's policyholders if the insurer is unable to fully pay their claims.

We develop a model that examines an insurance company's incentives to increase risk (relative to the social optimum) through its asset and underwriting quality choices. We demonstrate, in a world with asymmetric information, that the presence of a guaranty fund exacerbates such risk-taking by the insurer. We discuss certain regulatory provisions, which comprise of contingent payments that the insurer must make to the guarantee fund *ex post* in solvent states of the world that may alleviate such risk-taking incentives. However, it should be noted that a particular insurer's specific risk-taking incentives of course depend on that company's facts and circumstances and may thus be more complex than the incentives we have considered in our theoretical model. Our analysis should therefore be viewed as exploratory one intended to prompt further research into the optimal regulatory design of insurance guarantee schemes, which is an important policy issue in many countries.

Acknowledgments Sealey and Gandar acknowledge the financial support of the Financial Services Exchange, Charlotte, N.C. and the Belk College of Business, University of North Carolina at Charlotte. Earlier drafts on which this chapter is based benefited from comments of the participants of the Symposium on Insurance: Mathematics and Economics, and the American Risk and Insurance Association. The authors thank an anonymous referee and the editor, Dr. Malabika Roy, for their helpful comments.

Appendix

Proof of Proposition 1 First, consider the optimal asset quality decision of the social planner, q_A^{FB}, compared to that of the insurer's equity holders, q_A^{SB}. A necessary and sufficient condition for $q_A^{FB} > q_A^{SB}$ is that the RHS of (3) is less than the RHS of (6) when evaluated at the first-best solution. Simplifying (3) and (6) yields the following necessary and sufficient inequality

$$-(1-q_L)(L_2-R_2)\frac{[\xi^E(q_A^{FB}, q_L) - \xi^{SP}(q_A^{FB}, q_L)]P'_{q_A}(q_A^{FB}, q_L)}{P(q_A^{FB}, q_L)} < 0, \tag{18}$$

$\forall\, q_L \in (0,1)$. Recognizing that $P'_{q_A}(q_A^{FB}, q_L) > 0$, $(1-q_L)(L_2-R_2) > 0$, if $\xi^E(q_A^{FB}, q_L) > \xi^{SP}(q_A^{FB}, q_L)$, then the condition holds. It can be readily shown that $\xi^E(q_A^{FB}, q_L) > \xi^{SP}(q_A^{FB}, q_L)$ if $(1-q_A)(1-q_L)(L_2-R_2) > 0$, which is true from the specifications of the model. Note that (18) holds $\forall\, q_L \in (0,1)$.

Turning now to the differences in optimal underwriting quality for the social planner versus that of equity holders, a comparison of Eqs. (4) and (7) indicates that the necessary and sufficient condition for $q_L^{FB} > q_L^{SB}$ is the following inequality, evaluated at the first-best underwriting solution, holds

$$-(1-q_L)(L_2-R_2) + \frac{[\xi^E(q_A, q_L^{FB}) - \xi^{SP}(q_A, q_L^{FB})]P'_{q_L}(q_A, q_L^{FB})}{P(q_A, q_L^{FB})} < 0, \tag{19}$$

$\forall\, q_A \in (0,1)$. Similar to the discussion of (18), $(1-q_L)(L_2-R_2) > 0$ and $\xi^E(q_A, q_L^{SB}) > \xi^{SP}(q_A, q_L^{SB})$. Moreover, as was shown in an earlier section, $P'_{q_L}(q_A, q_L^{SB}) < 0$. Thus, the inequality in (19) holds, and $\forall\, q_A \in (0,1)$.

Q.E.D.

Proof of Proposition 2 To determine the impact of the subsidy on optimal decisions, first consider the asset quality decision. A comparison of the first-order conditions in (6) and (10) is required. Simplifying, if $q_A^{GF} < q_A^{SB}$, this requires that the following inequality holds when evaluated at the second best solution, q_A^{SB}:

$$\frac{q_A^{SB}(R(q_A^{SB}) - L_2) - q_L(L(q_L) - R_2) - q_A^{SB} q_L(R_2 - L_2)]P'_{q_A}(q_A^{SB}, q_L)}{P(q_A^{SB}, q_L)} > 0. \tag{20}$$

which follows from the fact that $R(q_A^{SB}) > L_2,\ L(q_L) < R_2,\ R_2 < L_2,\ P'_{q_A}(q_A^{SB}, q_L) > 0.$

Turning now to optimal underwriting decisions, if $q_L^{GF} < q_L^{SB}$, then the following inequality must hold when comparing (7) and (11):

$$\begin{aligned} &q_A(R(q_A) - L_2) + q_L(R_2 - L(q_L)q_A q_L(L_2 - R_2))] \\ &\times \left[\frac{(P^{GF}(q_A,\, q_L^{SB})P'_{q_L}(q_A,\, q_L^{SB}) - P(q_A,\, q_L^{SB})P^{GF'}_{q_L}(q_A,\, q_L^{SB}))}{P(q_A,\, q_L^{SB})P^{GF}(q_A,\, q_L^{SB})}\right] > 0. \end{aligned} \tag{21}$$

Substituting for $P^{GF}(q_A,\, q_L^{SB})$, $P'_{q_L}(q_A,\, q_L^{SB})$, $P(q_A,\, q_L^{SB})$, and $P^{GF'}_{q_L}(q_A,\, q_L^{SB})$, and simplifying yields

$$\begin{aligned} &q_A(R(q_A) - L_2) + q_L(R_2 - L(q_L)) + q_A q_L(L_2 - R_2))] \\ &\times \left[\frac{(L-R)[(L(q_L) - q_A L(q_L)) + (q_L L'(q_L) - q_L^2 L'(q_L)) - (q_A q_L L'(q_L) - q_A q_L^2 L'(q_L))]}{P(q_A,\, q_L^{SB})P^{GF}(q_A,\, q_L^{SB})}\right] > 0. \end{aligned} \tag{22}$$

The first term in brackets on the LHS in (14) was shown earlier to be positive. The numerator in the second term is positive, as is the denominator.

Q.E.D.

References

Benston GJ, Eisenbeis R, Horvitz P, Kane E, Kaufman G (1986) Perspectives on safe and sound banking: past, present, and future. MIT Press, Cambridge

Chan Y-S, Greenbaum SI, Thakor AV (1992) Is fairly priced deposit insurance possible? J Financ 47:227–245

Chen AH, Ju N, Mazumdar S, Verma A (2006) Correlated default risks and deposit insurance pricing. J Money Credit Bank 38(2):375–398

Demirgüç-Kunt A, Kane E, Laeven L (2014) Deposit insurance database. NBER Working Paper No. 20278. July 2014

Downs D, Sommers D (1999) Monitoring, ownership, and risk-taking: the impact of guaranty funds. J Risk Insur 66:477–497

Giammarino RM, Lewis TR, Sappington D (1993) An incentive approach to banking regulation. J Financ 48(4):1523–1542

Green R (1984) Investment Incentives, Debt and Warrants. J Financ Econ 13:115–136

John K, John TA, Senbet L (1991) Risk-shifting incentives of depository institutions: a new perspective on federal deposit insurance reform. J Bank Financ 15:895–915

Kuras G (2013) Insurance guarantee schemes: international practices and applicability to Russia. *Oliver Wyman* presentation, 7 November 2013. http://www.insur-life.ru/2013/files/presentations/Kuras.pdf. Accessed 26 April 2015

Lee S-J, Mayers D, Smith C (1997) Guaranty funds and risk-taking: evidence from the insurance industry. J Financ Econ 44:3–24

Ligon JA, Thistle PD (2008) Moral hazard and background risk in competitive insurance markets. Economica 75:700–709

Mazumdar S, Yoon S (1996) Loan monitoring, competition, and socially optimal bank capital regulations. J Risk Insur 63(2):279–312

Mazumdar S (1997) Regulatory monitoring, closure costs and bank moral hazard behavior. J Regul Econ 12(3):267–289

Nagarajan S, Sealey CW (1995) Forbearance, deposit insurance pricing, and incentive compatible bank regulation. J Bank Financ 19:1109–1130

Nagarajan S, Sealey CW (1998) State-contingent bank regulation and the fair pricing of deposit insurance. J Bank Financ 22:1139–1156

Yoon S, Mazumdar S (1996) Fairly priced deposit insurance, incentive compatible regulations and bank asset choices. Geneva Papers Risk insur Theory 23:121–141

Performance of Aggregate Portfolios of Equity Mutual Funds: Skill or Luck?

Rama Seth and Kamran Quddus

Abstract Mutual funds have become an important means for individual investors to invest in the equity market. They allow investors to buy into a growing stock market as well as diversify their risk. The U.S. Securities and Exchange Commission defines a mutual fund as "a type of investment company that pools money from many investors and invests the money in stocks, bonds, money-market instruments, other securities, or even cash." Assets under management by mutual funds total $30 trillion globally, or about 40 % of world's GDP. In developed markets like the United States, assets under management account for approximately 89 % of GDP. By contrast emerging markets like India remain highly under penetrated. The chapter will first present some descriptive statistics contrasting the two countries. It will then evaluate the performance of mutual funds in an emerging market such as India, borrowing a methodology extensively used in asset pricing literature (Fama and French, J Finance 318 Econ 33(1):3–56, 2010). Using bootstrap simulations, we will analyze the persistence of fund returns which will be used in distinguishing skill from luck.

1 Introduction

There are costs associated with active management of mutual funds that drive down the returns accruing to investors. Mutual funds charge sales load and other fees for provision of services like fund administration, portfolio management, etc. While some of these charges are paid indirectly by investors, others like entry and exit loads are charged as one-time expense either at sale or redemption of mutual fund units.

Fama and French (2010) examine mutual fund performance from the perspective of equilibrium accounting and note that after taking fund management expenses into account, the performance of actively managed U.S. equity mutual funds lag the passive benchmarks. They propose that because of the constraint on active investing,

R. Seth (✉) · K. Quddus
Indian Institute of Management Calcutta, Kolkata, India
e-mail: rama_seth@iimcal.ac.in

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_27

if the preexpense gross return for the active funds produces α; (abnormal expected return) close to zero, the same will be negative by the amount of fund expenses when estimated after adjusting for fund expenses. Using bootstrap simulation on the mutual fund returns, they compare the actual fund α to that obtained from simulations. In so doing true α is set to zero so that the distribution of estimated α reflects a population with no abnormal performance in fund returns. A comparison of simulated α to that of actual α, facilitates inference regarding the existence of skilled managers.

Following Fama and French (2010), our study investigates the performance of Indian mutual funds from the equilibrium accounting perspective. The study demonstrates that if the gross returns for the value-weighted portfolio of funds is close to zero, the same would be negative after taking fund expenses into account. Using bootstrap simulations, we analyze the persistence of fund returns, distinguishing skill from luck.

2 Understanding Mutual Funds

A *Mutual Fund* is a type of investment entity that pools money from investors and invests them in a wide variety of securities. The U.S. Securities and Exchange Commission defines a mutual fund as "a type of investment company that pools money from many investors and invests the money in stocks, bonds, money-market instruments, other securities, or even cash."

Worldwide, the mutual funds market consists of \$30 trillion[1] in assets under management and the United States alone holds a 50 % share with assets growing at a compounded annual growth of 13.6 % per annum (see Fig. 1).

Going by the number of funds (see Fig. 2), Europe has the maximum number of funds, while Africa has the least. But a closer look at the number of funds over a 7-year period from 2006 till 2013, reveals that the number of funds has grown at a maximum rate in United States (52.1 %), followed by Africa (41.6 %) and Asia (36.3 %). During the same period, growth in the number of funds worldwide was 23.2 % (refer Fig. 3).

Table 1 presents the region wise yearly figures for the number of mutual funds across the world for the period 2006–2013.

Variance in average annual mutual fund returns across the BRICS countries is not very much: 0.7–2 %. While Brazil offered the highest average monthly mutual fund return of 2.0 % over the sample period from October 2001 till October 2014, it also showed the widest range between the minimum and maximum achieved: −25 to 38 %. Indian funds had the highest median return of 2.2 % per month during the same period (see Table 2).

The types of funds chosen by US managers differ from those chosen by Indian managers. The majority of U.S. mutual fund assets are in long-term funds. Equity

[1]As per 2014 Investment Company Fact Book.

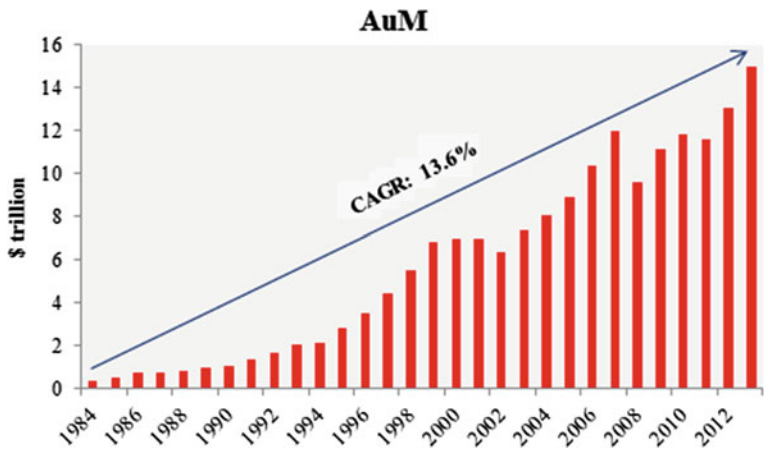


Fig. 1 Growth in assets under management in the US mutual fund market. *Source* ICI Research and Statistics

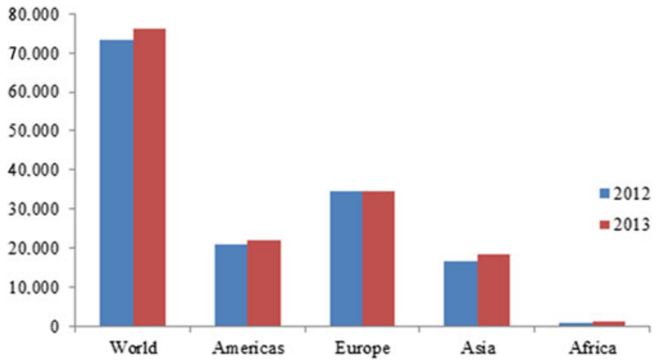


Fig. 2 Worldwide number of mutual funds. *Source* International Investment Funds Association

funds made up 52 % of U.S. mutual fund assets at year-end 2013 (Fig. 4). Domestic equity funds, or those funds that invest primarily in shares of U.S. corporations, held 38 % of total industry assets. World equity funds, or those that invest primarily in non-U.S. corporations, accounted for another 14 %. Bond funds accounted for 22 % of U.S. mutual fund assets, while money-market funds (18 %) and hybrid funds (8 %) held the remainder."[2]

The mutual fund composition of the mutual funds in U.S. reveals that equity funds comprised around 45 % of the total AuM at the end of December 2012.[3]

[2]A Review of Trends and Activities in the U.S. Investment Company Industry, ICI Fact Book, 2014.

[3]A Review of Trends and Activities in the U.S. Investment Company Industry, ICI Fact Book, 2013.

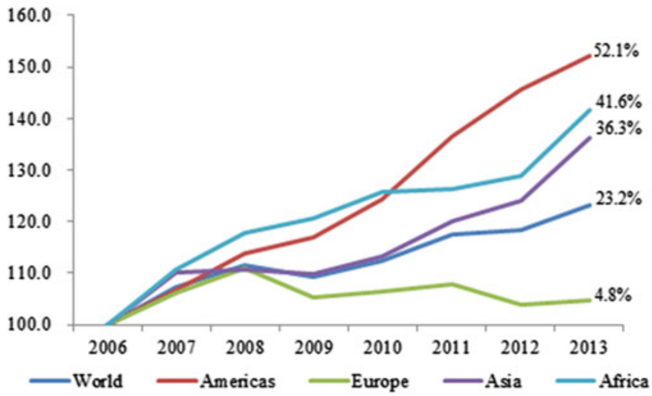


Fig. 3 Worldwide growth in the number of funds

Approximately 33 % were invested in domestic mutual funds in U.S. and nearly 12 % went into world equities comprising majorly non-U.S. mutual funds. The corresponding figures for Bond funds and money-market funds were 26 and 21 %, respectively.

The Indian mutual fund industry has a different composition and outlook. As Fig. 5 reveals, the majority of the mutual fund schemes in India are in the debt category (66.3 %), while equity schemes only constitute 14.5 % of the total assets under management.

As Fig. 6 depicts, average assets under management (AUM) grew at a CAGR of 21.1 % over the period September 2002 to September 2014. Average assets under management grew by 23 % for the year ended March 2013. This was considerably higher than the 12 % growth reported in March 2012.[4]

But in India, the share of equity mutual funds is much less compared to that of the US and it decreased 5 % between 2009 and 2013. By contrast, the debt segment grew significantly: 36 % over the same period. This difference in growth suggests that investors are still wary of investing in the market and are looking for relatively safer investments by directing their investments into fixed income instruments. Assets under management in the liquid and money market, and gold exchange traded funds (ETFs) grew by 16 and 18 %, respectively[5] (see Fig. 7) during the same period. Most recently investment in equity, debt, and money-market instruments have all shown an increase in 2014.

[4]Numbers obtained from AMFI, site (http://www.amfiindia.com/research-information/aum-data/average-aum).

[5]Based on Liquid/Money market AuM figure of Rs. 93,713 crores and Rs. 80,049 crores for Mar-2013 and Mar-2012 respectively. Corresponding figures for Gold ETFs are Rs. 11,648 crores and Rs. 9,886 crores, respectively. All data sourced through AMFI website.

Table 1 Worldwide number of mutual funds

Year-end	2006	2007	2008	2009	2010	2011	2012	2013
World	*61,854*	*66,344*	*69,029*	*67,526*	*69,486*	*72,600*	*73,229*	*76,200*
Americas	*14,474*	*15,456*	*16,456*	*16,928*	*17,986*	*19,742*	*21,089*	*22,020*
Argentina	223	241	253	252	254	281	291	297
Brazil	2,907	3,381	4,169	4,744	5,618	6,513	7,468	8,072
Canada	1,764	2,038	2,015	2,075	2,117	2,655	2,866	2,963
Chile	926	1,260	1,484	1,691	1,912	2,150	2,286	2,385
Costa Rica	100	93	85	64	68	63	66	66
Mexico	437	420	431	407	434	464	488	487
Trinidad and Tobago	NA	NA	NA	36	35	36	42	43
United States	8,117	8,023	8,019	7,659	7,548	7,580	7,582	7,707
Europe	*33,151*	*35,210*	*36,780*	*34,899*	*35,292*	*35,713*	*34,470*	*34,743*
Austria	948	1,070	1,065	1,016	1,016	1,003	995	981
Belgium	1,549	1,655	1,828	1,845	1,797	1,723	1,529	1,432
Bulgaria	NA	NA	81	85	90	92	95	98
Czech Republic	52	66	76	78	80	80	80	85
Denmark	494	500	489	483	490	500	495	510
Finland	376	379	389	377	366	368	375	369
France	8,092	8,243	8,301	7,982	7,791	7,744	7,392	7,154
Germany	1,199	1,462	1,675	2,067	2,106	2,051	2,059	2,012
Greece	247	230	239	210	213	196	177	166
Hungary	161	212	270	264	276	152	167	182
Ireland	2,531	2,898	3,097	2,721	2,899	3,085	3,167	3,345
Italy	989	924	742	675	650	659	600	661
Liechtenstein	233	391	335	348	409	437	535	657
Luxembourg	7,919	8,782	9,351	9,017	9,353	9,462	9,435	9,500
Malta	NA	NA	NA	NA	NA	59	54	69
Netherlands	473	450	458	NA	NA	495	497	501
Norway	524	511	530	487	507	507	406	573
Poland	157	188	210	208	214	226	259	264
Portugal	175	180	184	171	171	173	157	153
Romania	32	41	52	51	56	105	62	64
Russia	358	533	528	480	462	472	NA	NA
Slovakia	43	54	56	54	58	63	58	54
Slovenia	96	106	125	125	130	137	131	114
Spain	3,235	2,940	2,944	2,588	2,486	2,474	2,349	2,267
Sweden	474	477	508	506	504	508	456	484
Switzerland	609	567	572	509	653	664	667	765
Turkey	282	294	304	286	311	337	351	373
United Kingdom	1,903	2,057	2,371	2,266	2,204	1,941	1,922	1,910

(continued)

Table 1 (continued)

Year-end	2006	2007	2008	2009	2010	2011	2012	2013
Asia and Pacific	*13,479*	*14,847*	*14,909*	*14,795*	*15,265*	*16,198*	*16,703*	*18,375*
Australia	NA	NA	NA	NA	NA	NA	NA	NA
China	NA	341	429	547	660	831	1,065	1,415
Hong Kong	1,099	1,162	NA	NA	NA	NA	NA	NA
India	468	555	551	590	658	680	692	699
Japan	2,753	2,997	3,333	3,656	3,905	4,196	4,384	4,922
Korea, Rep. of	8,030	8,609	9,384	8,703	8,687	9,064	9,121	9,876
New Zealand	613	623	643	702	700	709	700	694
Pakistan	31	64	83	96	125	137	139	152
Philippines	38	40	43	41	43	47	48	47
Taiwan	447	456	443	460	487	534	554	570
Africa	*750*	*831*	*884*	*904*	*943*	*947*	*967*	*1,062*
South Africa	750	831	884	904	943	947	967	1,062

Source International Investment Funds Association

Table 2 Summary statistics for mutual fund monthly returns for BRICS countries

	Brazil	Russia	India	China	South Africa
Mean	2.0 %	1.4 %	1.9 %	0.7 %	0.7 %
Standard Error	0.0057	0.0055	0.0052	0.0048	0.0033
Median	1.5 %	1.9 %	2.2 %	0.9 %	0.9 %
Standard Deviation	0.0710	0.0693	0.0647	0.0607	0.0413
Sample Variance	0.0050	0.0048	0.0042	0.0037	0.0017
Kurtosis	4.5467	4.9035	1.9432	1.3517	3.6025
Skewness	0.6758	−1.0878	−0.3288	−0.3286	0.8196
Range	0.6250	0.5437	0.4895	0.3758	0.2866
Minimum	−24.6 %	−34.5 %	−22.1 %	−19.5 %	−8.7 %
Maximum	37.9 %	19.9 %	26.8 %	18.0 %	20.0 %
Confidence level (95.0 %)	0.0112	0.010929266	0.0102	0.0096	0.0065

Source Bloomberg Database

There is a lot of untapped potential for the industry. The main problem faced by the industry is the low level of financial inclusion leaving room for further penetration.[6] Business of Indian mutual funds remains concentrated within Tier 1 cities, and the industry is trying to enhance penetration ratio by reaching to customers in other cities.

[6]See "Untapped potential in Indian mutual fund market" CII-PwC Report—Mutual Fund Summit 2013, on Confederation of Indian Industry site. The top five cities of India constitute 74 % of the mutual fund business, with the remaining 26 % coming from all other cities.

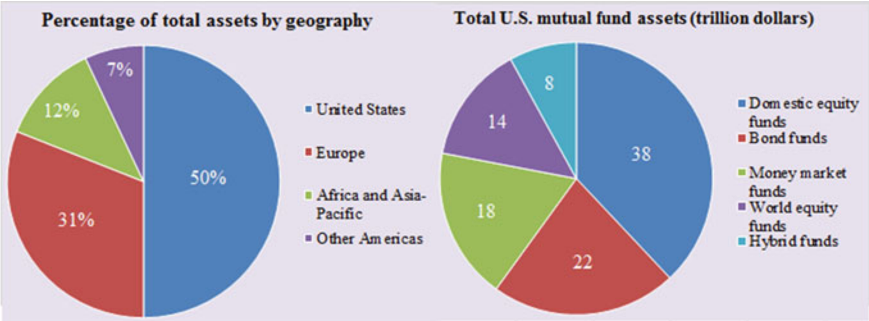


Fig. 4 Classification of US mutual fund market. *Source* ICI Research and Statistics

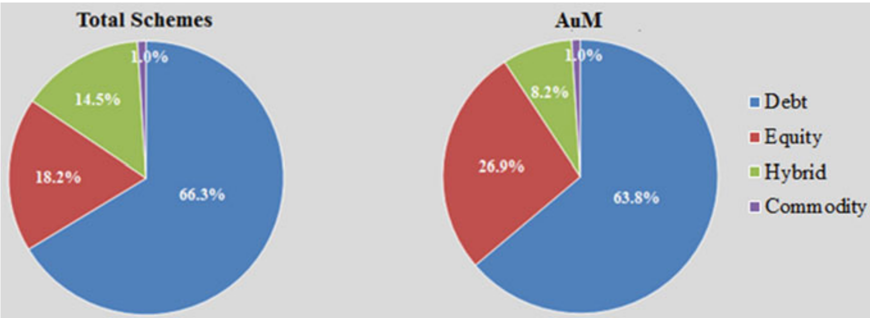


Fig. 5 Classification of Indian mutual fund market. *Source* Association of Mutual Funds of India

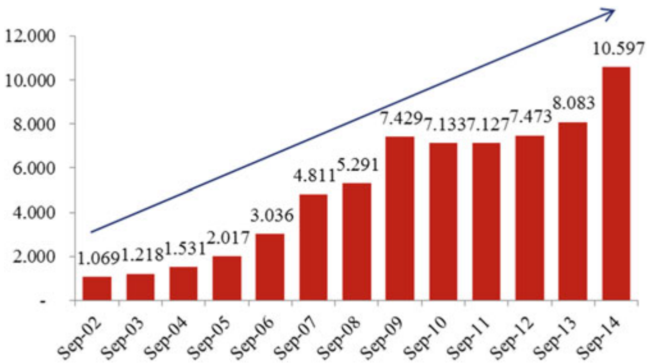


Fig. 6 Average asset under management (In bn INR)

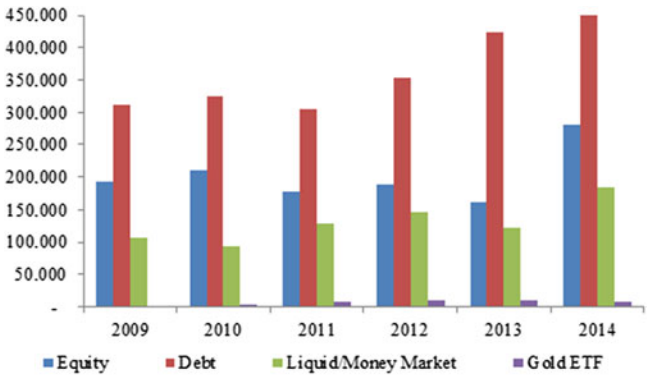


Fig. 7 Category-wise growth in assets under management

Though the scope of the industry is huge, challenges to avail of this scope exists owing very low levels of awareness and financial literacy. Even where the ability to invest exists, these savings are not channelized into mutual fund products. The funds are not invested in these instruments partly due to demand and partly due to supply factors. The relatively slow growth in capital markets restricts supply and lack of awareness of mutual funds being a potential investment vehicle restricts supply.

Cultural and behavioral factors further limit the demand for mutual funds. Indians still feel that gold and property are less risky assets as compared to capital market instruments. The reduction in risk through pooling of individual instruments that mutual funds offer, is not a feature easily understood by the lay investor. Better information dissemination and increased transparency of fees and return structure would likely improve the desirability of these instruments to the Indian investor. Strengthening distribution networks and enhancing levels of investor education in rural areas would likely result in large increases in investor demand.

In short, Indian markets differ from more developed markets like the United States. The Indian mutual fund industry is in early stages of development compared to its highly matured US counterpart. US mutual funds invest in a wide range of stocks spanning almost whole spectrum of the market. Indian mutual funds restrict their investment to only the highly liquid stocks. These differences have been highlighted by various researchers. Ramasamy and Yeung (2003) posit that although the number of mutual funds in emerging markets like India is small, they exhibit higher growth vis a vis their US counterparts. The study also highlights some of the peculiarities associated with emerging markets in terms of high volatility, extent of regulations[7] and quantum of government intervention.

[7]Emerging markets are characterized by higher regulatory shield, in terms of more closeness of their markets from global financial markets, etc.

2.1 Literature Survey

Prior studies have mostly revolved around evaluating the fund performance using Treynor, Sharpe and Jensen measures as the criteria for evaluating the performance of funds. These studies focus on individual fund-wise analysis, rather than taking an aggregate view. These studies also do not address whether abnormal returns would persist in the future or were just chance occurrences.

Roy and Ghosh (2012) examine the performance of gilt mutual funds during the recent recessionary period from January 2008 till February 2009. They study the performance of 31 open-ended gilt mutual funds, using the Sharpe and Treynor measures and find that Only 3 out of 10 public sector mutual fund schemes generates abnormal returns as indicated by positive and statistically significant alphas. Similarly, 7 out of 13 private sector and 6 out of 8 foreign mutual fund schemes generate abnormal returns. The market timing ability of fund managers is measured through γ_i, in the model specified below.

$$\mathrm{R}_i - \mathrm{R}_f = \alpha_i + \beta_i\left(\mathrm{R}_m - \mathrm{R}_f\right) + \gamma_i\left(\mathrm{R}_m - \mathrm{R}_f\right)^2 + e_i$$

The abnormal return from timing of market arises when γ_is are statistically significant. The study indicates that none of the gilt mutual funds generate any abnormal returns by timing[8] the market during the recession period.

Santhi and Gurunathan (2012) use Sharpe, Treynor, and Jensen risk adjusted measures to analyze the performance of 32 open-ended ELSS[9] mutual fund schemes for the sample period of 2006–2012. The study reveals that 5 schemes produced more than 2 % monthly average return, 11 schemes achieved more than 1 % rate of return, while the remaining 16 attained less than 1 % rate of return. The study reveals an interesting[10] finding that some of the mutual funds exhibited higher volatility that the stock market itself.

Rakesh (2012) shows that out of 28 diversified equity mutual funds, about 60 % were able to beat the market benchmark during the period ranging from January

[8]Jensen (1968) describes market timing as the manager's ability to predict the general behavior of security prices and accordingly engage in market timing by changing the portfolio betas to take advantage of market movements.

[9]ELSS refers to Equity Linked Savings Scheme that is meant for indirect participation in the stock market and provides tax benefits under Section 80(C) of the IT Act.

[10]Mutual Funds returns indicate some funds were able to beat the NSE and BSE benchmarks which generated an average monthly return of approximately 1.7 %, while the higher volatility documented in the study is surprising given the fact the mutual funds operate as a collective investment vehicle, which should diversify away the idiosyncratic risk components. However, the higher volatility evidence may be indicative of market timing on the part of managers who might be shifting the portfolio to higher beta and hence more aggressive stocks.

2007 to June 2011. Approximately, 58 % of the schemes generated superior returns by virtue of identifying the correct stocks and the fund managers failed to correctly time the market.[11]

Chakraborty et al. (2008) reveal that 23 of the 40 equity mutual funds under study have generated above normal returns compared to the benchmark. The study, however, does not test the skills of the fund managers.

Vijayakumar and Muruganandan (2012) investigate the fund of fund (FoF)[12] performance for different fund characteristics such as fund's volatility, average AuM, turnover ratio, income ratio and expense ratio. Using panel data analysis on a sample of 14 FoFs, the study analyzes their performance over 5 year period from 2004 to 2008. They find that the return generated by fund managers is not sufficient to cover the costs associated with the investments.

2.2 The Performance of Aggregate Portfolios of Indian Equity Mutual Funds

We have collected a sample of 63 mutual funds from the ACE Mutual Fund database, covering only those funds that invest majorly[13] in Indian common stocks.

2.3 The Regression Framework

Fama and French (1993) use a three-factor model, incorporating SMB and HML as additional factors over and above the market excess return to explain the excess return generated by the market. SMB and HML capture the premium captured by smaller market capitalization companies and high book-to-market companies, respectively. Carhart (1997) incorporated an additional factor WML that captures the premium associated with the price momentum.

We use the three-factor model of Fama and French (1993), but we also show results for Carhart's (1997) four-factor model. Our regression model is shown below.

[11]Jensen (1968) documents that a manager's superior forecasting ability may arise from either predicting the price movements of individual securities (security selection) or forecasting the general trend of market as a whole (market timing). Rakesh (2012) finds evidence in favor of stock selection skills, rather than market timing ability.

[12]Brands and Gallagher (2005) define Fund of Fund (FoF) as single managed investment portfolio, where the assets are invested across a number of individual managed funds.

[13]Mutual Funds in our sample set invest primarily in Indian equities, besides a small amount of corpus that may have been invested in other avenues, like parked in higly liquid short term mutual funds, etc.

$$R_{it} - R_{ft} = \alpha_i + \beta_i(\mathrm{MRKT}_t) + \gamma_i \mathrm{SMB}_t + \delta_i \mathrm{HML}_t + \lambda_i \mathrm{WML}_t + e_{\mathrm{it}} \tag{1}$$

R_{it} Return on fund i for month t,
R_{ft} Market return (the return on S&P CNX Nifty Index),
SMB_t Size factor of Fama and French (1993) model,
HML_t Value-growth returns Fama and French (1993) model,
WML_t Momentum return of the Carhart's (1997) model,
α_i Average return left unexplained by the benchmark model and
e_i Regression residual

As mentioned previously, SMB and HML relate to the Fama and French (1993) size premium and value premium respectively, whereas WML relates to the Carhart (1997) momentum factor. MRKT represents the excess return generated on the market portfolio.

2.4 Construction of Risk Factors

Consistent with the literature (Carhart 1997; Fama and French 1993) we construct four determinants as follows: market (MRKT, i.e., market return in excess of risk free rate of interest), represents the premium for undertaking market risk, while the other determinants are, size (SMB, i.e., small minus big), book-to-market equity or value (HML, i.e., high minus low) and momentum (WML, i.e., winners minus losers). To estimate excess market returns we use 91-day Treasury bill rate provided by the Reserve Bank of India as proxy for risk free rate (R_f) and value-weighted S&P CNX Nifty index returns provided by the National Stock Exchange of India as proxy for the market portfolio. Firm size (SZ) is proxied by the market capitalization (stock price times outstanding shares) at the end of each month. Book-to-market equity (BM) at the end of each month is the ratio of book value at the end of the latest quarter to the market value of equity at the end of that month. Momentum is the cumulative return of a stock in the preceding 12 months.

To construct SMB, HML, WML we consider 1 month holding return and at the end of each month, portfolios are rebalanced to calculate returns. Following Fama and French (1993), the six portfolios used for SMB and HML construction are BV (Big-Value), SV (Small-Value), BN (Big-Neutral), SG (Small-Growth), SN (Small-Neutral) and BG (Big-Growth) from the intersection of two SZ and three BM groups. SMB is measured each month as the equal-weighted average of the returns on the three small stock portfolios minus the return on the three big-stock portfolios, i.e. ((SG − BG) + (SN − BN) + (SV − BV))/3 which can be rewritten as ((SG + SN + SV) − (BG + BN + BV))/3. Similarly, HML is measured each month as the equal-weighted average of the returns on two high BM portfolios

Table 3 Summary statistics for monthly explanatory returns for the three-factor and four-factor models

	Average return	Standard deviation	t-statistic
MRKT	0.0044	0.0758	−0.5748
SMB	0.0043	0.0356	1.1886
HML	0.0100	0.0342	2.8926
WML	0.0848	0.0627	13.3154

Note Sample period consists of 97 monthly observations from May 2006 to May 2014. t-statistic reported is the ratio of mean to its standard error

minus the returns on the two low BM portfolios, i.e., ((SV − BV) + (BV − BG))/2 which can be rewritten as ((SV + BV) − (BV + BG))/2. In order to measure WML, we follow Carhart (1997) and construct four equal-weighted portfolios with the intersection of two size and two return momentum groups, i.e., WS (Winner-Small), WB (Winner-Big), LS (Loser-Small) and LB (Loser-Big). WML is the equal-weighted average of the returns on the two winner stock portfolios minus the return on the two loser stock portfolios, i.e., ((WS − LS) + (WB − LB))/2 which can be rewritten as ((WS + WB) − (LS + LB))/2.

Descriptive statistics in Table 3, reveal that MRKT, SMB, HML, and WML coefficients are all positive, but only HML and WML coefficients are statistically significant, as evidenced by high magnitudes of t-coefficients.

2.5 *Regression Results for Equal-Weighted and Value-Weighted Portfolios of Active Funds*

The Equal-weighted[14] returns are based on average return of mutual funds, with every fund receiving equal weighting of one, in the portfolio. Whereas, fund returns are weighted by their respective AuM in the value-weighted returns. In each case, the returns are shown on gross as well as net of mutual funds expenses. Net return, the return received by investors is calculated as

$$\text{Net Return} = (\text{Ending NAV} - \text{Beginning NAV})/\text{Beginning NAV}$$

Monthly expense ratio is added back to Net Return to arrive at Gross Return.

HML and WML slopes are close to zero indicating minimal exposure[15] of the value-growth and momentum factors of the mutual funds as a whole. This also

[14]Equal-weighted returns shows no bias towards funds with higher AuMs, whereas in value-weighted returns, larger funds with higher AuMs receive more weightage.

[15]This indicates that HML and WML factors are not able to explain the excess returns better as compared to other factors in the model.

Table 4 Regression results for Equal-Weight (EW) and Value-Weight (VW) Portfolios

	12*α						
	Net	Gross	MRKT	SMB	HML	WML	R^2
Panel A: EW returns							
Coef	−0.07	0.07	0.82				0.95
t(Coef)	−26.50	4.20	44.23				
Coef	−0.07	0.07	0.83	0.19	−0.07		0.96
t(Coef)	−26.74	4.38	47.62	4.99	−1.65		
Coef	−0.07	0.03	0.84	0.19	−0.06	0.03	0.96
t(Coef)	−14.88	1.22	45.01	5.06	−1.52	1.50	
Panel B: VW returns							
Coef	−0.07	0.05	0.51				0.94
t(Coef)	−28.32	4.24	39.47				
Coef	−0.07	0.00	0.51	0.03	−0.04		0.95
t(Coef)	−27.90	4.26	40.36	3.68	−1.30		
Coef	−0.06	0.07	0.51	0.05	−0.04	0.05	0.95
t(Coef)	−15.06	3.23	37.05	3.65	−1.38	−0.99	

implies that, manager's excess return cannot be attributed to value investing, i.e., investing majorly in low price to book ratio stocks, or momentum investing, i.e., chasing stocks that have been outperforming the market. SMB slope for equal-weighted returns, 0.19 is smaller than SMB slope for value-weighted returns, 0.03, implying a tendency of smaller funds to invest in smaller stocks. However, smaller SMB slope for value-weighted returns discredits the above tendency of smaller funds to invest in smaller stocks (refer Table 4).

2.6 Bootstrap Methodology

Bootstrap simulation technique is commonly used for estimating the reliability or accuracy of the forecast statistics.[16] In empirical simulation-based study, it is used to answer precision-based questions on the simulated data. Alternatively, they can also be used for testing of hypothesis.

First Monte Carlo simulation is used to generate an empirical estimate of the statistics sampling distribution, in our case, abnormal risk adjusted returns. The bootstrap is then used to distinguish fund managers who are "lucky" from those

[16]Mun, Johnathan. *Modeling Risk*. Hoboken, N.J.: Wiley, 2010, p. 159.

who are "skilled." We distinguish between α, the population mean, which we are trying to infer from the sample estimate $\ddot{\alpha}$ and estimate from bootstrap sample, $\tilde{\alpha}$.

If abnormal risk adjusted return excess return, i.e., $\alpha > 0$ (outperformance is attributed to skill of fund manager). We start by estimating $\acute{\alpha}$ from the sample. As a next step, data is resampled, creating a new time series using bootstrapping technique. The new time series constructed each time, constrains α, the true abnormal return to zero.

2.7 Bootstrap Analysis

Three-factor regressions indicated positive $\acute{\alpha}$ for nineteen funds, based on gross returns calculated. To further investigate whether the excess risk adjusted returns is due to the skill of the fund manager or just chance (sheer good luck), we further simulate the fund returns for those funds with positive $\acute{\alpha}$ (see Fig. 8).

Based on three-factor regression, the null hypothesis, Ho: $\alpha = 0$, gets rejected for all but one fund (refer Table 5) in favor of the alternative H1: $\alpha > 0$. That implies that out of 63 mutual funds that we started with, eighteen funds have managers with skills to generate superior performance.

We next examine whether managers of those 18 funds have sufficient skill to generate excess returns to cover the fund management costs. In other words, are the investors benefitted from the skills of those managers, in terms of extra returns getting passed-on to them? To answer the later question, the above analysis is repeated but this time with net returns.

The estimations of $\hat{\alpha}$ for the above shortlisted 18 funds are shown in the Table 6 below along with their critical values in the next column. Wherever $\hat{\alpha} - q$ (95 %) > 0, the null hypothesis is rejected in favor of the alternative H1. The table

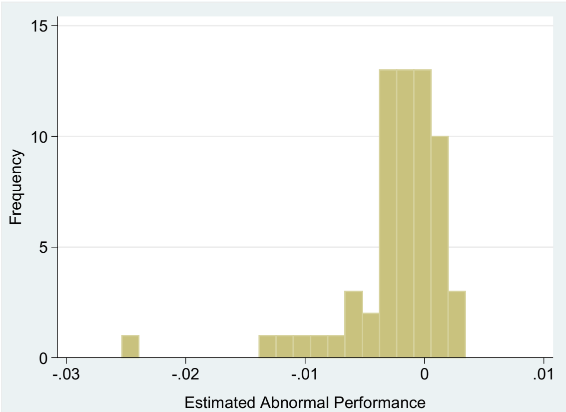


Fig. 8 Distribution of abnormal fund performance

Table 5 Critical values gross returns

Fund name	Gross return	
	$\hat{\alpha}_i$	$\hat{\alpha}_i - q(95\ \%)$
BNP Paribas Equity Fund-Direct Plan-Growth Option	0.00342	0.01793
Birla Sun Life India Gennext Fund-Growth Option	0.00258	0.01050
Axis Equity Fund-Growth	0.00232	0.00747
ICICI Prudential Focused Bluechip Equity Fund-Regular Plan-Growth	0.00171	0.01047
Edelweiss Diversified Growth Equity Top 100 Fund-Growth option	0.00167	0.00806
IDFC Nifty Fund-Regular Plan-Growth	0.00121	0.00285
Birla Sun Life Top 100 Fund-Growth Option	0.00100	0.00560
ICICI Prudential Long-Term-Regular Plan-Growth	0.00080	0.00333
IDFC Equity Fund-Plan B-Growth	0.00080	−0.00744
Canara Robeco Floating Rate-Regular Plan-Growth	0.00078	0.00136
Reliance Index Fund-Nifty Plan-Direct Plan Growth Plan-Growth	0.00065	0.00315
Religare Invesco Business Leaders Fund-Growth	0.00060	0.00421
Reliance Focused Large Cap Fund-Direct Plan Growth Plan-Growth	0.00059	0.00808
ICICI Prudential Dynamic-Regular Plan-Growth	0.00035	0.00916
LIC NOMURA MF Index Fund-Nifty-Direct Plan Dividend Option	0.00033	0.00169
Birla Sun Life Index Fund-Growth-Direct Plan	0.00030	0.00170
Taurus Nifty Index Fund-Direct Plan-Growth Option	0.00022	0.00138
BOI AXA Tax Advantage Fund-Regular Plan-Growth	0.00006	0.01111
UTI-NIFTY Index Fund-Growth Option-Direct	0.00006	0.00098

Table 6 Critical values net returns

Fund name	Net returns	
	$\hat{\alpha}$	$\hat{\alpha} - q(95\ \%)$
ICICI Prudential Focused Bluechip Equity Fund-Growth	0.01087	0.00489
BNP Paribas Equity Fund-Direct Plan-Growth Option	0.01066	0.00089
Axis Equity Fund-Growth	0.00880	−0.30249
Birla Sun Life India Gennext Fund-Growth Option	0.00838	0.00304
ICICI Prudential Dynamic-Regular Plan-Growth	0.00825	0.00325
Reliance Index Fund-Nifty Plan-Direct Plan Growth Plan-Growth	0.00789	0.00672
Reliance Focused Large Cap Fund-Direct Plan Growth Plan-Growth	0.00782	0.00335
Edelweiss Diversified Growth Equity Top 100 Fund-Growth option	0.00773	0.00291
LIC NOMURA MF Index Fund-Nifty-Direct Plan Dividend Option	0.00757	0.00679
Birla Sun Life Index Fund-Growth-Direct Plan	0.00754	0.00649
Taurus Nifty Index Fund-Direct Plan-Growth Option	0.00745	0.00673
UTI-NIFTY Index Fund-Growth Option-Direct	0.00729	0.00672
ICICI Prudential Long-Term-Regular Plan-Growth	0.00729	0.00551
IDFC Equity Fund-Plan B-Growth	0.00690	0.00510
Birla Sun Life Top 100 Fund-Growth Option	0.00680	0.00345
Religare Invesco Business Leaders Fund-Growth	0.00673	0.00448
Canara Robeco Floating Rate-Regular Plan-GROWTH	0.00658	0.00545
BOI AXA Tax Advantage Fund-Regular Plan-Growth	0.00600	0.00067

clearly shows that in all but one fund, the null hypothesis gets rejected. Hence, in 17 of the 63 funds that we started with, fund managers have sufficient skill to cover the extra expenses incurred in management of fund.

3 Conclusion

With the explosion of new mutual fund schemes in India, not all the funds are able to pass on the benefits to the investors. Even though the funds might be generating excess returns, it gets eaten away by the management itself.

This study throws interesting insights for those investors who have no time or expertise to participate directly in the market. Even while taking the mutual fund route, investors are not able to figure out which funds to invest in. Just going with the good performers may be tricky, as they might start underperforming after some time.

Using a sample of 63 mutual funds, the study shows that two were lucky enough to generate superior returns, but 17 of them were skilled enough to reward the investors for investing in those funds.[17] Further, the study shows in a systematic way, that these are skilled managers who are capable of generating benchmark adjusted excess returns,[18] even after covering their own internal fund management expenses. The results are consistent with Fama and French (2010) who find that after taking management expenses into account, the actively managed funds start lagging their passive counterparts.[19]

Our study has two limitations. First of all the period considered for the study is small, and second the sample set is limited to just the open-ended equity mutual funds[20] that invest primarily in the domestic equity market. Both limitations are a result of data constraints.

Going forward, the performance of funds can be seen in the context of performance of the underlying broader equity market. In case of Indian mutual funds, the broader market seems to be looking up. Further market imperfections make it possible for fund managers to somehow select the right stocks at right time which is bit difficult in case of US markets.[21]

---

[17]This distinguishes skilled managers from those who just got lucky by holding on to good stocks by chance or otherwise. Our study suggests that funds generate superior returns primarily because of skill, but there are a few funds that just got it by luck.

[18]This means that fund managers can beat the broader market index such as S&P CNX Nifty, using their superior fund managing skills.

[19]Some mutual funds do beat the benchmark, when returns are measured on a gross basis, i.e., not taking management expenses into account. The excess returns generated get offset when management expenses are taken into consideration.

[20]Other types of mutual funds could be debt-oriented mutual funds, money-market liquid funds, Gilt Funds, Fund-of-Funds, etc.

[21]This suggests that it may be possible for Indian fund managers with superior stock selection and market timing skills to generate excess returns for the investors.

References

Brands S, Gallagher DR (2005) Portfolio selection, diversification and fund-of-funds: a note. Acc Finance 45(2):185–197

Carhart Mark M (1997) On persistence in mutual fund performance. J Finance 52(1):57–82

Chakraborty M, Jain PK, Kallianpur V (2008) Mutual fund performance: an evaluation of select growth funds in India. South Asian J Manag 15(4)

Fama EF, French KR (1993) Common risk factors in the returns on stocks and bonds. J Financ Econ 33(1):3–56

Fama EF, French KR (2010) Luck versus skill in the cross-section of mutual fund returns. J Finance 65(5):1915–1947

Jensen M (1968) The performance of mutual funds in the period 1945–1964. J Finance 23:389–416

Kumar R (2012) Market timing, selectivity and mutual fund performance: an empirical investigation of selective equity diversified schemes in India (October 3, 2012). IUP J Financ Econ X(1):62–84.

Ramasamy B, Yeung MCH (2003) Evaluating mutual funds in an emerging market: factors that matter to financial advisors. Int J Bank Mark 21(3):122–136

Roy S, Ghosh SK (2012) A comparative study of mutual fund performance during recession in India. Int J Financ Manag 2.2

Santhi NS, Gurunathan KB (2012) An analysis of risk-adjusted return on tax-saving mutual fund schemes in India. IUP J Financ Risk Manag 9(3):54–71

Vijayakumar N, Muruganandan S (2012) The relationship between fund performance and fund characteristics: evidence from India. IUP J Appl Finance 18(2):5–18

Foreign Bank Presence and Financial Development in Emerging Market and Developing Economies: An Empirical Investigation

Sasidaran Gopalan

Abstract This chapter empirically examines the relationship between foreign bank presence and financial development in a panel framework covering 57 emerging market and developing economies (EMDEs) over 1995–2009. Specifically, the chapter undertakes an empirical investigation of three inter-related questions: One, how does greater foreign bank presence affect domestic credit creation in EMDEs? Two, are there threshold levels of foreign bank presence associated with financial sector development? Three, is the relationship between foreign banks and financial development conditional on the presence of a threshold level of institutional environment? While we find a positive relationship between foreign banks and financial development in general, the empirical results suggest that a strong information environment tends to strengthen the beneficial impact that foreign banks can have on financial development.

Keywords Foreign bank entry · Financial development · Institutional thresholds

JEL Classification F21 · G00 · G21 · O16

1 Introduction

The series of crises in emerging market and developing economies (EMDEs) during the 1990s covering the entire spectrum of countries from Latin America to Asia provided a major impetus for these countries to open up their domestic markets to foreign banks (Gopalan and Rajan 2010). Notwithstanding the variations in the

---

This chapter is based on and builds upon ongoing joint work with Ramkishen S. Rajan. Selected relevant references include Rajan and Gopalan (2014), Gopalan (2015a, b).

S. Gopalan (✉)
Asia Competitiveness Institute, Lee Kuan Yew School of Public Policy,
National University of Singapore, Singapore, Singapore
e-mail: sppsdg@nus.edu.sg

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_28

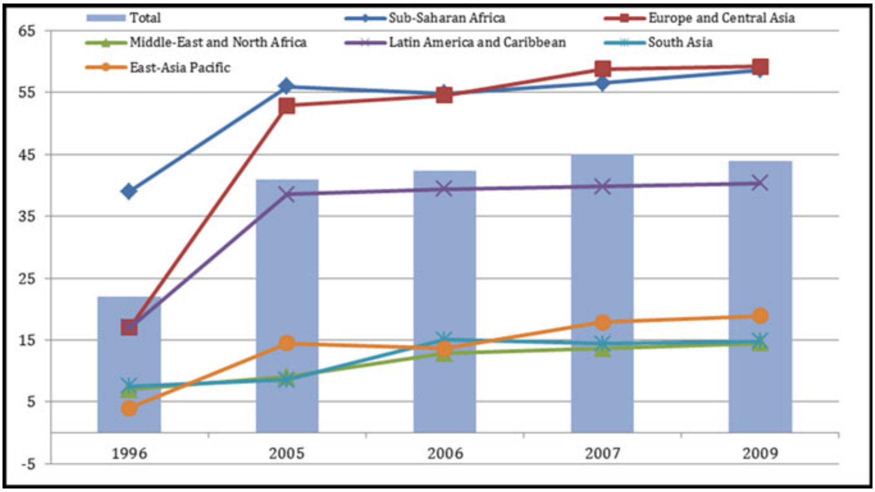


Fig. 1 Share of foreign bank assets in total banking assets across emerging markets and developing economies (EMDEs). *Note* This figure shows the average share of assets held by foreign banks (expressed as percentage of total assets) in each region at each point in time. A bank is considered foreign when it owns at least 50 % of shares. *Source* Compiled from Claessens et al. (2008), Claessens and Van Horen (2011)

degree and scope of their involvement between regions and countries since then, foreign bank presence has grown significantly across the board in the EMDEs. A preferable yardstick to measure the extent of foreign bank presence in a country is to look at the percentage share of their assets in the domestic banking system. Available data indicates that foreign banks in terms of their share of assets in the total banking system across EMDEs have doubled from 22 to 44 % between 1995 and 2009 (Fig. 1).[1]

There is a growing body of literature that indicates that foreign bank presence can bring in a variety of benefits to the host economy.[2] First, increased foreign bank presence can facilitate the process of international financial integration by plugging the host economies with international financial markets which allows them greater access to capital. Second, foreign banks contribute to the enhancement of the domestic banking regulatory and supervisory frameworks which improves the efficiency of the host banking system. Third, foreign bank presence helps reduce the

[1]At the outset, it is worth emphasizing that we would like to make a distinction between foreign bank *entry* and foreign bank *presence*. While technically much of the phenomenon we are describing relates to the effects of foreign bank *entry* and how it affects host countries through their *presence*, the metric we use in this chapter (foreign bank assets as a share of domestic banking assets) is a *stock* variable. Hence it does not capture *entry* (flow) in the strict sense of the term. Hence we use the term *foreign bank presence* instead of entry.

[2]See Gopalan (2015b), Rajan and Gopalan (2014) and references cited within for a discussion on the literature on foreign bank entry.

extent of "non-commercial" or "connected" lending as these banks are not as politically connected as the home-grown institutions and therefore less susceptible to political patronage. Fourth, foreign banks increase domestic banking competition and improve the quality of financial services, primarily through the new technologies and instruments they bring in and finally, foreign banks can also provide a safety valve given their diverse and comfortable liquidity sources.

Increasing foreign bank penetration in many EMDEs has given rise to a literature on its causes and consequences as well as debates. One such debate in the literature pertains to the contribution of foreign bank presence to overall financial sector development (Van Horen 2013). This chapter is interested in empirically investigating the relationship between foreign bank presence and financial sector development in selected EMDEs. The remainder of the chapter is structured as follows: Sect. 2 provides an overview of what financial sector development is and why the relationship between foreign bank presence and financial development is important. Section 3 briefly surveys the relevant empirical literature. Section 4 outlines the research hypotheses, the empirical methodology, the relevant data, and associated definitions. The empirical results are discussed in Sects. 5 and 6 concludes.

2 How Do Foreign Banks Affect Financial Development?

Financial sector development is a very broad term capturing "the factors, policies, and institutions that lead to effective financial intermediation and markets, and deep and broad access to capital and financial services" (World Economic Forum 2011, p. xiii). The literature relating to foreign banks specifically focuses on two aspects of financial development namely financial depth and financial inclusion, which in total are taken to contribute to overall financial sector development. In theory, foreign banks primarily through their credit creation could contribute to the deepening of the domestic banking system as commonly measured by expanding banking credit to private sector; they could also widen access to banking and financial services to a large segment of the population by promoting financial inclusion.

A handful of papers in the literature have found that foreign banks contribute to reduced costs of financial intermediation resulting in increased credit availability that in turn facilitates overall financial development in the host country (Claessens et al. 2001; Martinez Peria and Mody 2004). However, most other papers appear to be more sanguine about the positive impacts of foreign bank presence, especially in EMDEs. For instance Rashid (2011) finds that increased foreign bank presence leads to higher interest rate spreads, lower levels of private sector credit that in turn results in relatively weaker financial development. Other notable studies such as Claessens and Van Horen (2013), Detragiache et al. (2008) find that on average there is a *negative* relationship between foreign bank presence and financial development as measured by private credit creation, though the negative relationship could be largely driven by host country institutional differences.

In similar vein, there are also concerns on the financial inclusion front that the foreign bank presence could negatively impact banking sector outreach as captured for instance by a general decline in the number of deposit and loan accounts, owing to the tendency of foreign banks to cater to a smaller segment of the population (Beck and Martinez Peria 2009).[3]

Considering that banks play an important role in EMDEs in achieving financial inclusion and financial deepening where bank-based financial systems dominate other forms of providers of financial services, the question of how foreign banks affect financial sector development assumes policy significance. However, it must be emphasized that the relevant literature specifically investigating the relationship between foreign bank presence and financial sector development (as measured through credit creation) is not only limited but also ambiguous at best.

A parallel literature that warrants mention here consists of a set of studies that have investigated a much broader question of how openness to international financial flows affects financial development and thereby economic growth (for instance, see Baltagi et al. 2009; Calderon and Kubota 2009; Chinn and Ito 2006; Kose et al. 2011; Mishkin 2009). One of the central conclusions of this broader literature is that higher international financial openness is positively associated with domestic financial sector development, though this relationship is conditional on countries achieving a certain threshold level of institutional development. As the literature surveyed by Kose et al. (2011) examining the relationship between international financial openness and economic growth suggests, financial sector development is one of the primary "collateral benefits" of international financial openness and that such indirect benefits as the development of domestic financial markets and improving corporate and public governance contribute positively to overall economic growth in EMDEs.

Given that foreign bank entry/presence is one form of international banking liberalization—pertaining to 'commercial presence' in the General Agreement on Trade in Services (GATS) parlance—it may appear reasonable for one to expect that the entry of foreign banks will typically contribute to greater financial sector development in EMDEs. However, drawing on the findings from the broader literature on international financial flows, it becomes important to understand if the relationship between foreign bank presence and financial development is also governed by nonlinearities, i.e., do countries need to have a certain threshold of institutional development and/or foreign bank presence before experiencing desirable impacts a la financial development?

The possibility of the existence of threshold effects in terms of foreign bank presence and how it affects financial sector development definitely deserves further scrutiny especially given that the some papers like Detragiache et al. (2008), Rashid (2011) find a *negative* relationship governing foreign bank presence and financial

[3]This issue is closely related to the literature on the implications of foreign bank presence on firms' access to finance which has primarily focused on the dynamics of foreign bank lending to small and opaque firms (mostly the small-and medium-sized enterprises).

development. Further, Claessens and Van Horen (2013) also find that the negative result holds only for a subsample of developing countries and it disappears for emerging markets, suggesting that this relationship may not be necessarily linear. While possibly, the negative relationship need not indicate anything causal but instead could have been a result of the "non-random" entry of foreign banks into markets that were in crisis (Cull and Martinez Peria 2010), it is still unclear whether different degrees of foreign banking presence affects financial development differently.

In light of the dominant finding in the literature suggesting a negative relationship between foreign bank presence and financial development, the chapter is interested in asking the following set of inter-related questions: Are there thresholds of foreign bank ownership that one can identify beyond which a negative relationship with private credit creation turns positive? Second, related to this—is there also a possible existence of a threshold effect in terms of host country institutional development that is required for foreign banks to have a positive beneficial impact on financial sector development? While the literature points to such possibilities, a systematic examination of such threshold effects appears to be missing, a gap that this chapter attempts to fill.

We attempt to answer this question in two parts: First, we empirically test for the implications of greater foreign bank presence on domestic credit creation in EMDEs, including testing for foreign bank thresholds. In the second part, we test if the relationship is conditional on the presence of a threshold level of institutional environment? We survey the related empirical literature in greater detail in the next section before moving to the empirics of the chapter.

3 Related Empirical Literature

Based on the foregoing discussion, broadly, one may expect foreign banks to positively contribute to overall financial sector development by deepening financial markets and broadening access to financial services but the literature appears to be ambiguous. Not only are the results mixed, but the number of papers dealing specifically with the related issue is also surprisingly limited. But since there is a parallel literature that deals with the broader question of international financial openness and financial development, we briefly survey this literature first before specifically turning to the limited number of papers that look at the relationship between foreign banks and financial development.

A number of papers have tested for the relationship between international financial openness (broadly encompassing all types of capital flows) and financial development and have found that openness to international financial flows broadly serve as an important driver of domestic financial market development (Levine 1996; Chinn and Ito 2006; Baltagi et al. 2009; Calderon and Kubota 2009). However, as Chinn and Ito (2006) emphasize, the link between financial liberalization and financial development is not "unequivocal" as financial liberalization can have the desired impact on financial development only when the host economies, especially EMEs are equipped with some "reasonable" legal and institutional infrastructure. The rationale for the existence of such thresholds stems from the so-called "absorptive capacity" of these economies to internalize the benefits of such financial flows. For instance, Johnston et al. (1997) suggests that before a country's capital accounts are opened, the financial intermediaries need to be strengthened in order to guarantee the efficient use of capital inflows. Countries with weak financial systems may need time to develop financial institutions and markets, especially the banking sector, before liberalizing their capital account. Thus, the emphasis must be on establishing an effective system of prudential supervision before liberalizing the capital account. Along similar lines, with an emphasis on prudential supervision, Mishkin (2001), for instance argues that in order for financial liberalization to work and to avoid financial crises, institutional and governance prerequisites such as adequate prudential supervision as well as accounting and disclosure standards must be in place before opening up a country's capital account. Furthermore, a country characterized by weak legal infrastructure with ill-defined property rights may lack the capacity to strictly enforce contracts which could reduce the incentive for credit-related activities.

Thus the absence of legal protection for creditors and transparency in accounting rules could likely reduce the credibility of the domestic financial system, a point highlighted by the empirical study done by Chinn and Ito (2006). Upon examining the relationship between capital account openness and financial development proxied by stock market capitalization as a proportion of GDP—for a panel of 108 countries spanning 1980–2000, the authors find that that a higher level of financial openness (measured by *de jure* capital account openness index of Chinn-Ito) spurs equity market development only if a threshold level of legal development has been attained.

In a related paper, Baltagi et al. (2009) investigate whether the pace of financial development can be explained by the joint opening of both trade and financial sector openness. They use two datasets, one for 42 developing countries with banking credit to private sector as the dependent variable (1980–1996) and another for 32 industrial and developing countries with stock market capitalization as the proxy for financial development (1980–2003). They measure financial openness both using the *de jure* Chinn-Ito capital account openness index as well as the de facto ratio of foreign assets and liabilities to GDP sourced from Lane and Milesi-Ferreti (2007). Using these, the dynamic panel regressions the authors conduct suggest that trade and financial openness are statistically significant determinants of financial sector development.

In a similar study for an expanded sample, Calderon and Kubota (2009) test for the relationship between financial sector openness and domestic financial market development for a 145 economies from 1974 to 2007 and finds similar results. Specifically, they find that rising financial openness expands private credit, bank assets, and stock market and private bond market development and also generates overall efficiency gains in the domestic banking system. But consistent with other studies, the paper finds that the positive impacts are conditional on the level of institutional quality, the extent of investor protection, and the degree of trade openness. The measure of financial openness used is the ratio of foreign assets to GDP, foreign liabilities to GDP, and foreign assets and liabilities to GDP sourced from Lane and Milesi-Ferreti (2007) database.

By offering a comprehensive framework to examine such threshold conditions in the process of international financial integration and how it affects economic growth, Kose et al. (2011) unify all the disparate studies examining the relationship between different types of international financial openness and economic growth. Their empirical analysis tests for such threshold conditions for 84 countries between 1975 and 2004, by undertaking standard cross-country growth regressions. They use both parametric and semi-parametric estimations to capture threshold effects at the national level and also allow for nonlinearities in those threshold variables. Their findings suggest that there are "clearly identifiable thresholds" in key variables such as financial depth and institutional quality and that the growth benefits from financial openness significantly improves once countries pass those identified thresholds.[4]

The set of papers that specifically examine impact of foreign banks on financial sector development were discussed earlier in the previous section. Of the papers already discussed, two closely related papers to this chapter are Detragiache et al. (2008), Claessens and Van Horen (2013). Detragiache et al. (2008) use aggregate cross-country data for a sample of 89 lower-income countries to empirically test the association between foreign bank presence on growth in private credit levels proxied by commercial bank credit to the private sector. They test this using a standard cross-country regression framework as well as a dynamic panel framework, controlling for host country characteristics. The control variables they included in the model include GDP per capita measuring the overall levels of development in the country, inflation rates measured by the consumer price index (CPI) for each country, credit depth of information index capturing the cost to banks of obtaining information about borrowers, enforcement speed measuring the time taken to enforce a business contract, as well as a corruption index measuring the extent of corruption in each country. For the cross-sectional estimation they average the control variables over 1991–1998 and measure the dependent variable as a 3 year average over 1999–2002 while they use the data in its annual frequency for the dynamic panel estimation. Both their cross-section and panel estimation results suggest that the foreign bank participation is negatively associated with private sector credit.

As an extension, the authors hypothesize that if greater foreign bank presence is negatively associated with private credit levels, then they should observe a negative correlation between foreign presence and subsequent credit growth. Hence they examine the relationship between foreign bank presence and credit growth (as opposed to levels) and for the purposes of this estimation compute the dependent

[4]They also find that the thresholds are lower for foreign direct investment and portfolio equity liabilities compared to those for debt liabilities.

variable as the log difference of the private credit-to-GDP ratio in 1999–2001 and in 1994–1996. Their baseline specification similar to what was conducted in the first part of the analysis revealed consistent results in that they found foreign bank presence to have a negative and significant effect on credit growth.

Claessens and Van Horen (2013) perform a similar exercise for a sample of 111 countries representing all levels of development. While they in essence replicate the work of Detragiache et al. (2008), they do so for a larger sample of countries as well as use an updated database on foreign bank ownership which the authors themselves have compiled. They also test specifically for the relationship between foreign bank presence and private credit levels conditional on host country institutional characteristics similar to Detragiache et al. (2008) though they perform only the cross-country regressions in their paper as opposed to the variety of empirical specifications used by their source paper. They average the dependent variable—private credit to GDP ratio—over 2005–2007 along with a similar set of control variables as used by Detragiache et al. (2008), to include GDP per capita, inflation, the availability of information to creditors, and the time it takes to enforce contracts. They find that their results are broadly in line with those of Detragiache et al. (2008) but with one important qualification. Their negative relationship between private credit levels and foreign bank presence holds only in countries characterized by limited foreign bank presence coupled with costly access to information and contract enforcement. Further, they also find that the negative relationship holds only when the distance between the home country of the foreign bank and the host country is relatively "far." Interestingly, as Claessens and Van Horen (2013) observe, the negative relationship between foreign bank presence and private credit is only apparent for the subsample of developing countries though not very obvious for emerging markets. This heterogeneity in the results appears to underline, among other things, the importance of factoring in host country institutional characteristics.

While the two papers discussed above suggest a negative relationship between foreign bank presence and banking credit to private sector, it is inappropriate to infer these results as indicating the direction of causality. Furthermore, while both papers hint at the possibility of existence of threshold levels of institutional development and/or foreign bank presence to have a desired impact on financial development, neither explore the issue further in a systematic manner of identifying such thresholds, a gap that we attempt to fill in this chapter.

4 Research Hypotheses and Empirical Methodology

4.1 Key Hypotheses

Based on the foregoing discussion, we hypothesize that foreign bank presence affects financial development negatively up to a certain threshold beyond which the relationship turns positive. We also posit that greater information availability about potential creditors in an economy as well as strong creditor protection is significantly positively associated with increased credit creation by foreign banks.

4.2 Empirical Model

We proceed with our analysis in three steps. First, we specify a baseline panel regression following Detragiache et al. (2008) that helps us investigate whether the ratio of private credit to GDP (proxying for financial development) is correlated with the share of foreign bank assets over total banking assets in that country (proxying for foreign bank presence), controlling for a host of country-specific institutional factors. Second, we test for possible nonlinearities in the relationship by checking if foreign bank thresholds are important. Finally, as an extension, following Kose et al. (2011), we test for thresholds in terms of the conditioning variables that would be required for foreign bank presence to have an impact on financial development.

At the outset, it should be noted that the market share of foreign banks may be endogenous, as foreign banks could a priori choose to enter only those countries with a certain threshold of financial development. One plausible way to address this endogeneity problem is to use fixed-effects panel estimation as controls for unobserved country-specific fixed characteristics that might affect private credit growth. Hence, the chapter will estimate a fixed-effects panel data model, incorporating country (and time) fixed-effects.

We use cross-country panel data models to investigate potential thresholds in the relationship between foreign bank presence and financial development. We follow standard linear panel data specifications with the relationship between foreign bank presence and banking sector credit modeled as a nonlinear function using a quadratic specification. The second part of the analysis will test for the appropriate thresholds of the conditional variables which will be the host country institutional characteristics that will be interacted with the key foreign bank independent variable. We will focus on selected 57 emerging and developing economies based on Claessens and Van Horen (2011) for 1995–2009 for which we have information on foreign bank presence.

The basic estimating equation will be as follows:

$$y_{it} = \delta_i + \beta fb_{it} + \gamma \boldsymbol{X}_{it} + \mu_t + u_{it} \tag{1}$$

where,

y_{it} is the ratio of private credit to GDP of country i at time t;
δ_i is the country fixed effect;
fb_{it} is the share of bank assets held by foreign banks in country i at time t;
$\boldsymbol{X}_{it}$ is a matrix of control variables measured at time t;

μ_t	is the time fixed effect and
u_{it}	is the idiosyncratic error term;
β and γ	are the parameters to be estimated.

The parameter of interest is β which represents the coefficient of the foreign bank share. Based on the literature,[5] the following variables that could possibly affect private credit creation in an economy have been selected. We also discuss the priors that we expect to see from our empirical results below.

GDP per capita measuring the overall levels of development in the country and we expect a positive relationship between credit creation and countries with higher levels of development.

Inflation rate as measured by the consumer price index (CPI) for each country and we expect inflation to have an adverse impact on credit creation.

Exchange-rate regime exchange-rate regime of country *i* at year *t*; a priori, there is no particular direction of relationship that expect with respect to exchange rate regimes and credit creation.

Bank Z-Score capturing the probability of default of a country's banking system, calculated as a weighted average of the *Z*-scores of a country's individual banks (based on the individual banks' total assets). The *Z*-score compares a bank's buffers (capitalization and returns) with the volatility of those returns. We would expect a negative relationship between a higher *Z*-score and financial development as a higher probability of default of a country's banking system could likely lead to greater risk aversion that might hinder financial development.

Public Debt reflecting the gross public debt as a percentage of that country's GDP. We expect greater public indebtedness to lower credit creation, as greater pressures on authorities to persist with or introduce financial repression will likely hinder financial development.

Credit Depth of Information Index captures the cost to banks of obtaining information about borrowers and we expect that higher information availability to be positively associated with credit creation in the economy.

Legal Rights Index measuring the strength of legal rights in an economy that protect the rights of borrowers and lenders. We expect countries with stronger legal protection of creditors to have deeper credit markets and thus carry a positive sign.

Corruption Index a measure of the extent of corruption in each country that could also be used as a proxy to test the presence (or absence) of "politically connected" lending with greater foreign bank presence. We expect that countries with higher corruption to be associated with less credit to private sector in the presence of foreign banks.

Regulatory Quality Index measuring the quality of domestic regulations. We expect the variable to be positively related to financial development.

[5]For a detailed investigation on various possible determinants of private credit, see Djankov et al. (2007).

The nonlinear effect can be captured using a quadratic specification of the form:

$$y_{it} = \delta_i + \beta_1 fb_{it} + \beta_2 fb_{it}^2 + \gamma \boldsymbol{X}_{it} + u_{it} \quad (2)$$

As noted earlier, the presence of foreign banks could be endogenous as the levels of financial development in a country may determine the decision of a foreign bank to enter that market. Either they could choose to enter a country with low financial development because they consider that market to possess stronger growth prospects or to enter only those countries with higher levels of financial development. A priori, as Detragiache et al. (2008) point out, it is not clear how endogeneity might bias the coefficient β that represents foreign bank share. However, using fixed-effects estimation, we can control for unobserved country-specific fixed characteristics that might affect private credit levels.[6]

The second part of the analysis will be to include interaction terms involving foreign bank presence and the institutional variables that will help us test the importance of different institutional settings on the decision of credit creation of foreign banks in emerging economies. By doing this, we will be able to test for the impact of foreign bank presence change with defined thresholds of other conditional variables which could be qualify as prerequisites for banking liberalization to have a positive and desired impact on financial development. The methodology will follow Kose et al. (2011) in specifying both a linear interaction term between foreign banks and thresholds of conditioning variables as in (1), as well as nonlinear quadratic specifications as in (2).

$$y_{it} = \delta_i + \beta_1 fb_{it} + \beta_2 \mathrm{Inst}_{it} + \beta_3 fb_{it} * \mathrm{Inst}_{it} + u_{it} \quad (3)$$

This approach tests if the level of a particular institutional variable affects the marginal effect of foreign bank presence on financial development. The specification we employ implies that the marginal effect (either positive or negative) of foreign bank presence on financial development is larger at higher levels of the threshold institutional variable.

A quadratic interaction that allows for nonlinear effects of the threshold variable

$$y_{it} = \delta_i + \beta_1 fb_{it} + \beta_2 \mathrm{Inst}_{it} + \beta_3 fb_{it} * \mathrm{Inst}_{it} + \beta_4 \mathrm{Inst}_{it}^2 + \beta_5 fb_{it} * \mathrm{Inst}_{it}^2 + u_{it} \quad (4)$$

This allows for the possibility that, beyond a certain level, the threshold variable becomes more or less important in determining the marginal effect of foreign bank presence on financial development.

[6]Further, a key point to note here is that the estimates of the fixed-effects estimation remain robust only if the potential source of endogeneity arises from the correlation between the time-invariant component of the error term and the regressor of interest. This is because a fixed-effects model resolves this problem by excluding the unobservable time-invariant effects through a time-demeaning of the data.

4.3 Data Sources and Definitions

Data on foreign bank presence comes from a newly collected database put together by Claessens and Van Horen (2011) with the panel dataset containing information on bank ownership for 57 countries over the period 1995–2009. We focus on 57 countries in the sample that classify as emerging and developing economies as listed in Appendix Table 4. The data on all control variables will be sourced from World Bank World Development Indicators (WDI) and Global Financial Development Indicators (GFDI) database. All the control variables used in the empirics, their definitions and the sources are summarized in Appendix Table 5.

5 Empirical Results

We start with our baseline model as outlined in Eq. (1) that estimates the relationship between foreign bank presence and financial development controlling for macroeconomic, institutional, and financial variables.[7] The results of our panel estimation using country and time fixed effects with robust standard errors are summarized in Table 1. Column (1) in Table 1 reports estimates of our baseline without foreign bank presence. Column (2) reports the estimation results of our baseline model with the foreign bank variable.

As Column (1) shows, GDP per capita and creditor information turn out to be highly statistically significant determinants of financial development. The significance of GDP per capita is suggestive of the importance of accounting for heterogeneous levels of economic development while creditor information underlines the importance of a better information environment for enhancing financial development. It is interesting to note that bank *Z*-score, a measure of bank risk, appears to be only marginally significant at the 10 % level. The results suggest that the direction of the relationship for all the significant variables conforms to our priors.

In Column (2) we provide the estimates of our baseline specification including the foreign bank variable. Contrary to the findings of the previous literature, not only do our baseline results prove to be slightly better in terms of greater number of variables becoming significant, but we also find a positive relationship between foreign bank presence and financial development. GDP per capita continues to be highly significant at the 1 % level reiterating its positive association with greater financial development in an economy. The statistical significance of *Z*-score

---

[7] An examination of the correlations among the right-hand side variables rules out issues of multi-collinearity. The results are available on request.

Table 1 Fixed-effects estimates—impact of foreign bank and institutional thresholds on financial development

Dep var: Ln private credit to GDP (%)	(1) FB with inst thresholds (full sample)	(2) FB with inst thresholds (dev econ)	(3) FB with inst thresholds (EME)
Ln GDP per capita	1.432***	0.626**	2.336***
	(0.160)	(0.246)	(0.253)
Ln inflation rate	0.0634***	0.0185	0.0627***
	(0.0170)	(0.0255)	(0.0222)
Public debt (%)	−0.00151*	−0.00229*	0.00227*
	(0.000869)	(0.00131)	(0.00130)
Exchange-rate regime	−0.0322**	−0.0255	−0.0423**
	(0.0152)	(0.0292)	(0.0180)
Bank *Z*-score	−0.00779***	−0.00908	−0.00652**
	(0.00259)	(0.00653)	(0.00267)
Legal rights	−0.605***	0.585	−0.615***
	(0.138)	(0.778)	(0.142)
Creditor information	0.118***	0.251***	0.0562
	(0.0342)	(0.0566)	(0.0446)
Corruption	0.148*	−0.301*	0.222**
	(0.0786)	(0.166)	(0.102)
Foreign bank assets (%)	0.00394	0.0108***	0.000732
	(0.00258)	(0.00386)	(0.00340)
Foreign bank squared	−4.44e-05*	−0.000113***	−3.43e-05
	(2.31e-05)	(3.83e-05)	(2.81e-05)
Creditor information squared	−0.0249***	−0.0463***	−0.0129
	(0.00704)	(0.0106)	(0.00930)
Legal rights squared	0.0536***	−0.0516	0.0513***
	(0.0113)	(0.0669)	(0.0114)
Corruption squared	−0.144*	−0.577***	−0.122
	(0.0872)	(0.166)	(0.107)
Foreign bank* creditor Information squared	0.000193***	0.000250**	0.000173***
	(5.24e-05)	(0.000100)	(6.04e-05)
Foreign bank* legal rights squared	−4.42e-06	−2.96e-05	3.22e-05
	(3.19e-05)	(8.89e-05)	(3.84e-05)
Foreign bank* corruption squared	0.00218	0.00347	0.000140
	(0.00185)	(0.00276)	(0.00248)
Observations	436	163	273
R-squared	0.571	0.685	0.651
No of countries	54	21	33
Country fixed effects	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes

Notes ***Significant at the 1 % level; **Significant at the 5 % level; *Significant at the 10 % level; Robust standard errors clustered for countries in parenthesis indicates that an increase in the probability of banking default is negatively associated with financial development. Though lack of corruption is marginally significant, it is broadly indicative that better governance contributes positively to financial development. Further, we also find inflation to be statistically significant at the 5 % level. While the direction of relationships of all the other variables continue to be consistent with our priors, neither creditor information nor foreign bank variable are in accordance with our priors, with the creditor information variable turning insignificant, and foreign bank presence exhibiting a positive relationship—a departure from the earlier literature that found a negative relationship. However, it must be noted that the economic significance of foreign bank presence is quite marginal.

In order to verify if this association varies when we split our sample into developing and emerging economies with differing foreign bank presence, we run the same baseline model separately for developing and emerging economies. The results are summarized in columns 3 and 4. Interestingly, we observe that while foreign bank presence and income per capita continue to be significant and consistent with our baseline model, the significance of other variables is inconsistent with the full sample results. Moreover, the direction of relationship between foreign banks and financial development is consistently positive in both emerging and developing economies though the relationship is much stronger in developing economies relative to emerging economies. This seems to suggest that in relatively more developed economies[8] (emerging as compared to developing), the marginal impact of foreign bank presence appears to be decreasing, which is consistent with and corresponds to one of the important results of Gopalan (2015a). Further, this also seems to suggest that threshold levels of foreign bank presence might be important in assessing how they affect financial development.

To understand if thresholds of either foreign bank presence and/or institutional environment matter, we test for their importance. While we report the estimation of (2) in Table 2, the results for the regression specified in (3) and (4) are reported in Table 3. Starting with Table 2, we break the estimation into two parts. The first three columns of Table 2 provide the results of our specification in (2) for the full sample as well as developing and emerging economies subsamples, leaving out institutional controls. The last three columns repeat the estimation process with the institutional controls included. Going by the convention in the literature, we introduce a quadratic term of foreign bank presence to test if different levels of foreign bank presence have a different impact on financial development. We find the quadratic foreign bank term to be weakly significant at the 10 % level only for the emerging economies subsample and the direction of the coefficient appears to be

[8]The average income in our sample of 34 countries is US$4765 for emerging economies and US $1930 for our sample of 23 developing countries.

Table 2 Fixed-effects estimates—impact of foreign bank presence on financial development

Dep var: Ln private credit to GDP (%)	(1) Without FB	(2) With FB	(3) Developing econ	(4) Emerging econ
Ln GDP per capita	1.117***	1.537***	1.058**	2.474***
	(0.356)	(0.351)	(0.386)	(0.480)
Ln inflation rate	0.0238	0.0619**	0.0394	0.0550
	(0.0272)	(0.0281)	(0.0326)	(0.0352)
Public debt (%)	−0.00189	−0.00231	−0.00437*	0.00193
	(0.00177)	(0.00161)	(0.00217)	(0.00254)
Exchange-rate regime	−0.0199	−0.0212	−0.0229	−0.0283
	(0.0303)	(0.0243)	(0.0445)	(0.0221)
Bank *Z*-score	−0.00997*	−0.00848**	−0.00844	−0.00765**
	(0.00514)	(0.00374)	(0.00817)	(0.00301)
Legal rights	0.0489	0.0404	0.0277	0.0195
	(0.0695)	(0.0533)	(0.0537)	(0.0625)
Creditor information	0.0814***	0.0373	0.0385	0.0325
	(0.0276)	(0.0259)	(0.0443)	(0.0332)
Corruption	0.194	0.206*	0.157	0.261*
	(0.137)	(0.106)	(0.155)	(0.141)
Foreign bank assets (%)		0.00374**	0.00626***	0.00286*
		(0.00162)	(0.00217)	(0.00203)
Observations	516	436	163	273
R-squared	0.454	0.521	0.555	0.601
Number of countries	55	54	21	33
Country fixed-effects	Yes	Yes	YES	YES
Time fixed effects	Yes	Yes	YES	YES

Notes ***Significant at the 1 % level; **Significant at the 5 % level; *Significant at the 10 % level; Robust standard errors clustered for countries in parenthesis

in the opposite direction. Even though the results are contrary to our priors as well as the broad result emanating from the related literature, we believe that it is indicative of the point we reiterated earlier about the diminished marginal effects of foreign bank presence as countries move up the income ladder. We also find that broadly, the importance of other institutional control variables in terms of the direction of the coefficients and significance levels is broadly consistent with the results reported in Table 1.

The results of (3) and (4) as given in Table 3 underline two more interesting points. One, foreign bank presence appears to contribute positively and significantly to financial development only in developing countries (as given in Column 2 of Table 3). Two, there are clear institutional thresholds at work in order to see

Table 3 Fixed-effects estimates—impact of foreign bank thresholds on financial development

Dep var: Ln private credit to GDP (%)	(1)	(2)	(3)	(4)	(5)	(6)
	FB thresholds (full sample)	FB thresholds (dev econ)	FB thresholds (EME)	FB thresholds with inst (full sample)	FB thresholds with inst (dev econ)	FB thresholds with inst (EME)
Ln GDP per capita	1.728***	1.022***	2.738***	1.548***	0.806***	2.573***
	(0.125)	(0.158)	(0.194)	(0.138)	(0.191)	(0.208)
Ln inflation rate	0.00210	0.00682	−0.00577	0.00145	0.00375	−0.00532
	(0.0166)	(0.0238)	(0.0226)	(0.0165)	(0.0238)	(0.0226)
Public debt (%)	−0.00413***	−0.00492***	−0.00195	−0.00438***	−0.00538***	−0.00204*
	(0.000803)	(0.00111)	(0.00122)	−0.000796)	(0.00113)	(0.00121)
Exchange-rate regime	−0.0220	0.00205	−0.0457**	−0.0202	0.00454	−0.0444**
	(0.0156)	(0.0268)	(0.0192)	(0.0154)	(0.0268)	(0.0191)
Foreign bank assets (%)	0.00620***	0.0104***	0.00523*	0.00561***	0.0102***	0.00484*
	(0.00218)	(0.00313)	(0.00289)	(0.00216)	(0.00314)	(0.00288)
Foreign bank squared	−3.46e-05	−4.27e-05	−4.60e-05*	−2.80e-05	−4.87e-05	−3.85e-05
	(2.13e-05)	(3.23e-05)	(2.79e-05)	(2.12e-05)	(3.25e-05)	(2.79e-05)
Creditor information				0.0373*	0.0594*	0.0256
				(0.0196)	(0.0305)	(0.0260)
Legal rights				0.0880***	0.00862	0.0761***
				(0.0262)	(0.0577)	(0.0292)
Observations	646	242	404	645	241	404
R-squared	0.442	0.474	0.522	0.457	0.485	0.533
No of countries	55	22	34	54	21	34
Country fixed effects	Yes	Yes	yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	yes	Yes	Yes	Yes

Notes ***Significant at the 1 % level; **Significant at the 5 % level; *Significant at the 10 % level; Robust standard errors clustered for countries in parenthesis positive impacts of foreign bank presence on financial development. Specifically, they emphasize the importance of a strong creditor information environment and how foreign banks can positively contribute to financial development when countries have better informational environment, i.e., when the information environment is above a certain threshold.

6 Conclusion

Rising foreign bank participation in many emerging and developing economies (EMDEs) has given rise to a large body of work exploring its multifaceted impacts. One of the unsettled debates in the literature on foreign bank entry pertains to the contribution of foreign bank presence to financial sector development in the host economy. Foreign banks are expected to positively enhance financial sector development through their impact on credit creation and lowered cost of financial intermediation. They are also expected to contribute to greater equity and bond market liquidity thus enhancing financial development in an economy. However, there is growing contention that foreign bank entry results in raising interest rate spreads, lowering credit creation and negatively affecting financial development. In this context, this chapter has examined the relationship between foreign bank presence and financial development in 57 EMDEs, between 1995 and 2009 and also empirically tested if threshold levels of foreign banks and or institutions matter in this relationship.

The findings of the chapter appear to run counter to the existing literature in this field where our fixed-effects panel data estimation suggests that foreign bank presence is significantly and positively associated with financial development and is robust across different subsamples based on income classification.

While the chapter finds weak evidence for the existence of foreign bank thresholds in how they affect financial sector development, it does find strong evidence for the importance of institutional thresholds in order for foreign banks to have a beneficial impact on financial development. Specifically, the empirical results highlight that the positive relationship between foreign bank presence and financial development becomes stronger in countries with a certain threshold level of institutional development, especially in developing economies.

Appendix

See Tables 4 and 5.

Table 4 Full Sample—list of countries and regions

Region	Country
East Asia and Pacific (EAP)	Indonesia, Korea, Malaysia, Philippines, Thailand
Europe and Central Asia (ECA)	Armenia, Bulgaria, Croatia, Czech Republic, Estonia, Georgia, Kyrgyz Republic, Latvia, Lithuania, Moldova, Romania, Russia, Serbia, Slovak Republic, Slovenia, Ukraine.
Latin America and Caribbean (LAC)	Antigua, Argentina, Bolivia, Brazil, Chile, Colombia, Dominican Republic, Guatemala, Jamaica, Mexico, Panama, Paraguay, Peru, Uruguay, Venezuela
Middle East and North Africa (MENA)	Algeria, Bahrain, Jordan, Kuwait, Libya, Morocco and Oman
South Asia (SA)	India, Pakistan and Sri Lanka
Sub-Saharan Africa (SSA)	Madagascar, Mali, Mauritius, Mozambique, Namibia, Niger, Rwanda, Senegal, South Africa, Togo, Zimbabwe

Table 5 Sources and definitions

Variable	Definition	Source
Foreign bank assets (%)	Share of foreign bank assets in total banking assets	Claessens and Neeltje van Horen (2011), Claessens et al. (2008)
Inflation (average CPI: 2005 = 100)	Average inflation measured by consumer price index in 2005 prices	Global financial development database—world bank
GDP per capita (constant 2000 USD)	GDP per capita measured in 2000 US dollars	Global financial development database—world bank
Public debt	General gross government debt as a percentage of GDP	IMF historic public debt database
Exchange-rate regime	1. No separate legal tender/preannounced pegs 2. crawling pegs narrower than or equal to ±2 % 3. managed floating 4. freely floating 5. freely falling 6. dual market in which parallel market data is missing	Ilzetzki et al. (2008)
Private credit to GDP	The financial resources provided to the private sector by deposit money banks as a share of GDP. Deposit money banks comprise commercial banks and other financial institutions that accept transferable deposits, such as demand deposits. (International Monetary Fund, International Financial Statistics, and World Bank GDP estimates)	Global financial development database—world bank

(continued)

Table 5 (continued)

Variable	Definition	Source
Bank *Z*-score	Capturing the probability of default of a country's banking system, calculated as a weighted average of the *Z*-scores of a country's individual banks (based on the individual banks' total assets). The *Z*-score compares a bank's buffers (capitalization and returns) with the volatility of those returns	Global financial development database—world bank
Creditor information	This index measures rules and practices affecting the coverage, scope and accessibility of credit information available through either a public credit registry or a private credit bureau. (0 = low to 6 = high)	Doing business database—world bank
Corruption	Control of Corruption captures perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests. Estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from approximately −2.5 to 2.5	World governance indicators—world bank
Legal rights	This index measures the degree to which collateral and bankruptcy laws protect the rights of borrowers and lenders and thus facilitate lending	World development indicators—world bank

References

Baltagi B, Demetriades P, Law SH (2009) Financial development, openness and Institutions: Evidence from panel data. J Dev Econ 89:285–296

Beck T, Martinez Peria MS (2009) Foreign bank participation and outreach: Evidence from Mexico. J Financ Intermed 19:52–73

Calderon C, Kubota M (2009) Does financial openness lead to deeper domestic markets? World Bank policy research working paper 4973, June

Chinn MD, Ito H (2006) What matters for financial development? Capital controls, institutions and interactions. J Dev Econ 81:163–192

Claessens S, Van Horen N (2011) Foreign banks: trends, impact and financial stability. DNB working paper 330

Claessens S, Van Horen (2013) Foreign banks: trends and impact. J Money, Credit Banking. Supplement to 46(1):295–326 (2014)

Claessens S, Demirgüc-Kunt A, Huizinga H (2001) How does foreign entry affect domestic banking markets? J Bank Finance 25:891–911

Claessens S, Van Horen N, Gurcanlar T, Mercado J (2008) Foreign bank presence in developing countries 1995–2006: Data and Trends. Mimeo

Cull R, Martinez Peria MS (2010) Foreign bank participation in developing countries: what do we know about the drivers and consequences of this phenomenon? World bank policy research working paper 5398

Detragiache E, Tressel T, Gupta P (2008) Foreign banks in poor countries: theory and evidence. J Financ 63:2123–2160

Djankov S, McLiesh C, Shleifer A (2007) Private credit in 129 countries. J Financ Econ 84:299–329

Gopalan S (2015a) Does foreign bank entry contribute to financial depth? Examining the role of income thresholds? HKUST IEMS working paper 2015–05, Hong Kong University of Science and Technology

Gopalan S (2015b) Financial liberalization and foreign bank entry in emerging market and developing economies: what does the literature tell us? J Int Commer Econ Policy 6(2):1–25

Gopalan S, Rajan RS (2010) Financial sector de-regulation in emerging Asia: focus on foreign bank entry. J World Invest Trade 11(1):91–108

Ilzetzki E, Reinhart CM, Rogoff KS (2008) Exchange rate arrangements entering the 21st century: Which anchor will hold? May 13, Background Material. http://personal.lse.ac.uk/ilzetzki/

Johnston RB, Echeverria C, Darbar SM (1997) Sequencing capital account liberalization-lessons from the experiences in Chile, Indonesia, Korea, and Thailand (EPub), International Monetary Fund

Kose MA, Prasad E, Taylor AD (2011) Thresholds in the process of international financial integration. J Int Money Financ 30(1):147–179

Lane P, Milesi-Ferretti G (2007) The external wealth of nations mark II. J Int Econ 73: 223–250

Levine R (1996) Foreign banks, financial development, and economic growth. Int Financ Markets: Harmonization Versus Competition 7:224–254

Martinez Peria MS, Mody A (2004) How foreign participation and market concentration impact bank spreads: evidence from latin America. J Money, Credit, Banking 36(3):511–537

Mishkin FS (2001) Prudential supervision: why is it important and what are the issues? In: Prudential supervision: what works and what doesn't. University of Chicago Press, pp 1–30

Mishkin FS (2009) Globalization and financial development. J Dev Econ 89(2):164–169

Rajan RS, Gopalan S (2014) Economic management in a volatile environment: monetary and financial issues, Chap. 7–10. Palgrave-Macmillan, UK

Rashid H (2011) Credit to private sector, interest spread and volatility in credit-flows: Do bank ownership and deposits matter? United Nations Department of Economic and Social Affairs Working Paper No. 105

Van Horen N (2013) Foreign banks: access to finance and financial stability. In Acharya V, Beck T, Evanoff D, Kaufman G, Portes R (eds) The social value of the financial sector. Too big to fail or just too big? World Scientific Stud Int Econ 29:323–344

Banks, Financial Derivatives, and Crises: A Fourth-Generation Model

Romar Correa

Abstract Our fundamental is a national employment guarantee scheme. The regime is supported by the issue of government bonds with the Central Bank as purchaser of last resort working through the intermediation of commercial banks. Banks are also free to hold foreign bonds. The exchange rate adds an exogenous element to the accounting requirements. Finally, we consider the 'truth-telling' role that the Central Bank can play in dispelling manias and panics.

Keyword The government bond

1 Introduction

At this time of writing, recessionary conditions prevail in most countries in the world. Inflation is regarded as less of a problem. Fulfilling time-honoured expectations, governments are expected to step in and originate virtuous cycles of employment and activity. At the same time, the inefficiencies of government intervention are lodged in both professional and public consciousness. The task, in short, is to work out novel schemes to deliver socially desirable outcomes without treading familiar paths that lead to suboptimal scenarios. Thus, the budget deficit will continue to be regarded with concern. Since an increase in taxation is unlikely to be entertained anywhere, the consequence is a squeeze in government expenditure. At the same time, Central Banks (CBs hereafter), at the present moment, are freed from the constant pressure to calibrate the short-term interest rate so as to

I am grateful to an anonymous referee for his/her incisive criticisms of an earlier draft. Any errors and ambiguities that remain are entirely mine.

R. Correa (✉)
Department of Economics, University of Mumbai, Mumbai, India
e-mail: romarcorrea10@gmail.com

© Springer India 2016

M. Roy and S. Sinha Roy (eds.), *International Trade and International Finance*,
DOI 10.1007/978-81-322-2797-7_29

track the inflation rate. Finally, a macroeconomic description of any nation today must include its increasing embeddedness in real and financial flows to and from other nations. Our motivation lies in the fact that more and more countries of all types, ranging from Brazil to the European Union, are independently drawing up roadmaps for the future that contain common elements. In the first place, members of both the private and the public sector are sitting at tables together writing up long-term plans that explicitly eschew the short-termism of purely financial metrics. Investment Banks are being set up for the purpose under the stewardship of the CB. In a sophisticated reversion to the old planning literature, the job is to look inward first and then work out the appropriate open-economy adjuncts. In terms of our theme, we cannot afford to have a domestic employment guarantee scheme fall prey to a currency crisis. The literature distinguishes between first-, second-, third-generation models of currency crisis (Gandolfo 2002). We borrow elements from all three and take the next step into what we term a fourth generation model of crisis. The elements common to all the models are private agents making informed choices based on their anticipations of government positions on the exchange rate.

The familiar way to proceed would be to take a pick from the shelf of open-economy general equilibrium models. We prefer, instead, the structure of so-called Stock-Flow-Consistent (SFC) modelling. No inconsistency between the strategies is implied. The only discipline we impose is the double-entry book-keeping requirement of balance sheet accounting. Secondly, so-called 'stock-flow norms' are the drivers and brakes of difference equations. We deploy the classic Godley and Lavoie 2007, manual for the purpose. Our notations and definitions have been culled from the book. A mild advantage of the modelling style is that we do not have to contend with corner solutions like the impossible trinity. Indeed, under current conditions, an alternative 'irreconcilable duo' has been proposed (Bordo and James 2015). Independent monetary policy is possible only if the capital account is managed. The Mundell trilemma points in the direction of closer coordination, of policies that constrain domestic choices or capital controls as a way of stretching the national policy space. The optimal sequence is to build up domestic capacities before installing inducements to capital inflows. In the case to be discussed below, the employment of workers on domestic worksites is likely to be jeopardized in the event of large-scale cross-border movements of factors of production. The sharp implications of fixed versus floating foreign exchange rate regimes will not apply. In reality, all countries can be regarded as points along a continuum between the two ends, nowhere reaching the extremes of the textbook. Indeed, the connection between corner solutions and crises has been unravelling. In one panel data study of ninety developed and developing countries, crises, in the sense of interconnected banking, currency, and debt breakdowns, were not more intimately associated with regimes at one or other end of the spectrum (Combes et al. 2013). Intermediate regimes were more crises prone.

To illustrate, below is as stylized a balance sheet of a CB that will be found (Lavoie 2013).

Central Bank Balance Sheet

Assets	Liabilities
Forex reserves	Cash and reserves
Government Securities	
Advances to commercial banks and financial system	

The following reading does not depend on whether the economy is on a fixed or a flexible exchange rate system. With forex reserves acting as a buffer, CBs can set real or nominal interest rates. A fortiori, when countries run BOP surpluses CBs are not compelled to lower interest rates. The increase in that element in the balance sheet of the CB will be compensated for by a corresponding decrease in the other elements. With forex reserves acting as a residual, we can consider situations when advances to commercial banks are low but not just holding but operations in the government securities market are active. At the same time, the two items are connected. In a circle to be traced below, government securities are contracts that underpin the loan covenants that commercial banks enter into to support government projects.

2 From the Third to the Fourth Generation

The basis of the third-generation model is domestic investment. We transfer attention to the counterparty to this programme, the commercial bank that underwrites this investment. Our attention is drawn to the blueprints for failure-proof banking being written all over the world to support both employment-led growth as well as ensure that the financial entities that mediate between borrowers and lenders operate with built-in stabilizers. The 'leader of the club of banks', the CB, clearly has a pivotal role to play here. However, the time is long past when CBs could impose ceilings and floors on interest rates, even at 'home', or direct credit to sectors that maximize social welfare. The CB possesses Chartalist power but cannot ignore private sector incentives and arbitrage conditions. Finance was always fungible and therefore hard to contain within Chinese walls. Now, with financial innovation and the breaking down of other barriers, cross-border flows are impossible to stem. The requirement of a fourth-generation model of financial crises is to take cognisance of the changed nature of financial flows across nations since the 1980s (Forrest 2014). Growth in the industrialized world is sluggish while financial activity has increased. Large tranches of financial capital, at least a hundred times higher than the value of trade in real goods and services, prowl the world in search for high returns. Financial derivatives provide a leveraged platform for trading. Since the instruments are highly leveraged, speculators can trade in products that are valued at many multiples of the capital mobilized.

In our operation twist, we suggest that there is no reason why market-making government authorities comprising of the CB and the Treasury cannot write such derivative contracts. The particular government instrument that we will focus on exclusively is the long-term bond. Current scholarship on policy, as well, has shifted attention from the short to the long end of the yield curve. Quantitative easing in the USA has meant a burgeoning Fed balance sheet. From an international accounting perspective, the implication must be shrinking balance sheets of CBs that are linked to the Fed. There is an obvious tension here with the agenda of the 'home' CB that might be identically positioned to hold long-term government bonds to support the production of goods and services at home.

To simplify, in this section we consider a financial institution whose assets are funded by non-monetary liabilities which are bank-run proof. It is sufficient to motivate this assumption by reference to the Chicago Plan, the complementary aspect being introduced in the next section (Benes and Kumhof 2012). In other words, the liabilities cannot be near-monies. Beginning at home with a pared-down balance sheet of a commercial bank, we have

Commercial Bank Balance Sheet I

Assets	Liabilities
Loans	Own equity

Following Godley and Lavoie, we call e_b the number of equities supplied by the bank and p_{eb} the price of these equities. In that case, the item on the liabilities side of the balance sheet above is $e_b p_{eb}$. Loans, familiarly, are denoted by L. Having defined our stock balance sheet, we turn to flow elements. As indicated, in the current period a menu of private–public projects emerge on stream, leading to an increase in investment. That investment is supported by ΔL. We do not consider equity funding so as to focus on the elements of interest. Also, a massive employment guarantee scheme impacting on the entire country is likely to require the support of a network of banks rather than stock exchanges. Now, denoting the dividends of banks by FD_b and their profits by F_b, the following 'stock-flow norms' apply. The dividend yield of banks is defined as $r_K = FD_b/e_{b-1}p_{eb-1}$, where the -1 denotes the value of variables in the previous period and the price earnings ratio $PE = p_{eb}e_{b-1}/F_b$. With r_l as the rate of interest on bank loans, the balance sheet above can be equivalently expressed as follows. We distinguish between the current account and the capital account of the bank. The former is the flows of revenues and outlays that banks earn and disburse. Below, the only item under this head is the interest earned on loans made in the previous period (assuming one-period contracts). The capital account can be regarded as a balancing column showing how the change in their assets, loans, must have a counterpart in the change in their liabilities, equity.

Commercial Bank Balance Sheet II

Current	Capital
$r_{l\text{-}1}L_{\text{-}1}$	$-\Delta L$
	$+\Delta e_b p_{eb}$
0	0

Equating the two sides of this balance sheet and assuming that the dividend yield is unchanging in the current price and quantity of equity,

$$L = (1 - r_{l-1})L_{-1} + FD_b/r_K - PE \cdot F_b \tag{1}$$

The steady state or stationary solution of this difference equation is

$$r_l L + PE \cdot F_b = FD_b/r_K \tag{2}$$

Since two of the three terms in the expression contain ‘stock-flow norms’ r_K and PE, we are prompted to regard the loan rate in the third term also as a norm. It is a catchall for problems of adverse selection with the promoters of the projects as well as moral suasion from the CB. In the next section we will endogenize the loan rate making it coincide with the bond rate. Since government bonds are near-monies, the explanation is that money is ‘credit money’.

Now, suppose the commercial bank responds positively to the latest tranche of ‘foreign’ government bonds sale, BL^*, with p^*_{bL} as the price of the perpetuity. These variables without stars, to be introduced in the next section, are the domestic equivalents. Our domestic bank balance sheet extends to

Commercial Bank Balance Sheet III

Current	Capital
$r_{l-1}L_{-1}$	$-\Delta L$
	$+\Delta e_b p_{eb}$
	$-\Delta BL^* p^*_{bL}$
0	0

Clearly, with the costs of monitoring and auditing domestic business, the L item in the balance sheet might be dominated by returns to be earned in holding 'foreign' bonds. The loan item can drop off, the bank transmuting to a financial institution. Adding the term to Eq. 1, the modification is

$$L = (1 - r_{l-1})L_{-1} + FD_b/r_K - PE \cdot F_b - \Delta BL^* p^*_{bL} \tag{3}$$

Labouring the point, if the 'foreign' authorities are implementing a massive 'quantitative easing' exercise with an attendant 'low' interest rate regime, to that extent the fresh loan disbursement plans of the domestic bank are curtailed. Godley and Lavoie exploit the following elaboration of the new term in the spirit of calculus: $\Delta BL^* \Delta p^*_{bL} = \Delta BL^* p^*_{bL} + BL^* \Delta p^*_{bL}$. The term on the extreme right-hand side of this expression is capital gains, CAG. The fresh issue of bonds abroad might be for the purpose of financing a budget deficit. In the steady state, (budget balance), the ΔBL^* element would vanish. In addition, we multiply the 'foreign' terms by the exchange rate, xr. Rewriting Eq. 3 in consistent 'rupee' or 'dollar' terms,

$$L = (1 - r_{l-1})L_{-1} + FD_b/r_K - PE \cdot F_b - (\Delta BL^* \Delta p^*_{bL} - CAG)xr \tag{4}$$

Causes for concern, naturally, lie in the extreme right-hand side of the equation. The first term in brackets impacts directly on our theme. Under any interest parity conditions, its sign will determine outflows or inflows. Naturally, interest rate movements cut both ways: In one direction, they would swell the forex reserves of banks, in the other they could contribute to massive depletion of that component. In the steady state, the counterpart of Eq. 2 is

$$r_l L + \mathrm{PE} \cdot F_b = \mathrm{FD}_b/r_K + \mathrm{CAG} \cdot xr \tag{5}$$

Our remarks about stability extend to the product of the CAG and the exchange rate as well. It can be positive or negative, denoting gains or losses, and if the magnitude of forex bond holding is large, huge shrinkage or swelling of balance sheets would be the result. In the next section, we propose the domestic government bond as a stabilizing device.

3 Manias and Panics

Imagine a massive rural-cum-urban assembly of projects designed to generate jobs across the country. The blueprints are drawn up in the current period. When on stream they promise a generous rate of return α_2 some years from now. However, if the endeavour is terminated at any time before the terminal date, the result is catastrophic with a very low rate of return α_1 at any date before completion. We have introduced two states of the world, the 'good' state when the project fructifies and the 'bad' state with shells of unfinished work dotting the landscape. Our objective is to protect the network of employment-generating activity from fads and fashions that wrongly dump the economy into the bad state.

We propose that the domestic government bond, *B*, be a put option on the project exercised at time 1. The bond promises r_{l1} and r_{l2} over the current time period and the future state of the world. Recall that the rates are the loan rates of banks. The relationship is likely to reduce the uncertainty and information asymmetry that underlies standard private debt contracts. In brief,

$$B_1 = \max[r_{l1} - \alpha_1, 0]$$

From our description, clearly

$$\alpha_1 < r_{l1} < \mathrm{r}_{l2} < \alpha_2$$

The inequalities are a no-arbitrage condition.

We defer the next step of moving to risk-neutral probabilities for later. Indeed, we insist on holding on to the special traits of our menu of projects. If anything, an ambitious nationwide employment-generating programme that starts ab novo should not present any frequency distributions.

For our purposes, this means that we need to introduce one more financial asset. The government bond tracks the project and we have, in effect, one asset. With two states of the world, we are one short à la the requirement of complete markets. So, introduce an option that promises r_1^s in the 'bad' state and r_2^s in the 'good' state. That is, $r_1^s < r_2^s$.

We have two assets and two states. People will have preferences over asset bundles. The population is divided into two types. Let us call, arbitrarily, optimists (*O*), those who expect $\alpha_1/\alpha_2 < r_1^s/r_2^s$ and pessimists (*P*), those for whom the inequality is reversed. Now, financial institutions arrive on the scene to offer mutually beneficial arrangements to households who hold wealth *W*, part of which, *S*, is to be parked in the financial asset, the remainder, *K*, in the physical project. A contract must specify payouts $D_t(\theta)$ where *t* ranges over 'today' and 'tomorrow' and θ ranges over the two characters, *O* and *P*. One of the implicit constraints that must be fulfilled is the incentive-compatibility constraint. It should not benefit either party to pretend to be the other. The two truth-telling conditions are

$$\alpha_1 D_1(P) + \alpha_2 D_2(P) \leq \alpha_1 D_1(O) + \alpha_2 D_2(O)$$
$$r_1^s D_1(O) + r_2^s D_2(O) \leq r_1^s D_1(P) + r_2^s D_2(P)$$

It is straightforward to combine these inequalities to find that either one of two situations must hold, but not both. Either $\alpha_1/\alpha_2 \leq r_1^s/r_2^s$ or the other way around. That is, either the optimists have it or the pessimists call the shots.

Working backwards to the present, the distinction between the two types will vanish. If, arbitrarily, the population is pessimistic, finance leads and the real collapses.

Can a bank preserve the behavioural variety? Indeed, can we consider variegated financial institutions to that end? The supplement to the Chicago Plan introduced in the last section is the separation of the monetary and credit functions of the banking system. Accordingly, we can entertain a 100 % backing of deposits by fiat money. A riskless deposit contract would specify a repayment schedule in units of CB money in both states of the world. The deposit rates duple is (r_1^d, r_2^d). Suppose that the proportion of optimists in the population is μ. In that case, the total payout in the first period is $[\mu D_1(O) + (1 - \mu)D_1(P)]$. Likewise, the total payout in the second period is $[\mu D_2(O) + (1 - \mu)D_2(P)]$. It is not difficult to derive the arbitrage-free pairs of numbers that the commercial bank must offer. Technically, for mean-preserving returns we need $1 > r_1^d/r_2^d > \alpha_1/\alpha_2 > r_1^s/r_2^s$. The deposit contract satisfies both optimists and pessimists. The narrow bank rests on the first of the three inequalities. Unity is the return to central bank money and it nestles as close as possible to the return on bank deposits. The commercial bank of the earlier section would offer an equity contract that would specify interim payouts in the form of dividends in both states of the world. Dividends are not lump sum but are paid as a percentage of the price in the two periods. Let us call the reciprocal of the price of bank equity in both situations (r_1^{eb}, r_2^{eb}). They are known ex ante. In an identical fashion, we would derive the arbitrage-free pairs of numbers that the commercial bank must offer:

$$1 > r_1^{eb}/r_2^{eb} > \alpha_1/\alpha_2 > r_1^s/r_2^s.$$

We pick up the thread introduced with optimists and pessimists. The cycle informs our reasoning and, thus, optimists and pessimists are not given à priori. People are the first or the second depending on whether they are on a long upswing or in an unremitting downturn. Our treatment of these matters is drawn from the book by Chamley 2004. Nature chooses a state $\alpha_i \in$ A, with $i = 1, 2$. The two values can be normalized to 0 and 1, the bad state and the good state, respectively. Each agent receives a signal s that is informative on **A**. An agent uses the signal to update her prior distribution on **A**. Denote her prior density on the state of the project as $f(\alpha)$ and the distribution of s conditional on α by the density $B(s|\alpha)$. For our two states, Bayes' rule translates to the log likelihood ratio (LLR)

$$\log[f(\alpha_2|s)/f(\alpha_1|s)] = \log[f(\alpha_2)/f(\alpha_1)] + \log[B(s|\alpha_2)/B(s|\alpha_1)]$$

The updating term on the extreme right-hand side is independent of the individual LLR. We develop the representation of the term as a government bond-cum-updating measure.

The pdf of any agent is characterized by one number, the probability of the good state. The private signal of any individual takes the value of 1 or 0 with the following probabilities.

		Signal	
		$s = 1$	$s = 0$
state of project	α_2	q	$1 - q$
	α_1	$1 - q$	q

Agents with a signal $s = 1$ are optimists, those with a bad signal are pessimists. If the good signal is observed, the LR between the good and the bad states is updated by the updating multiplier $B(s = 1|\alpha_2)/B(s = 1|\alpha_1) = q/(1 - q)$. The number is greater than 1 iff $q > ½$. In that case, the signal $s = 1$ increases the probability of the good state. It is a good signal. If $q < ½$, the signal $s = 0$ would be the good signal. Here, $q < ½$ with $s = 1$ is a bad signal. We need a breakup of the LLR for our definitions. Thus,

$$\{\log[f(\alpha_2|s)\} - \log[f(\alpha_1|s)]\} = \{\log[f(\alpha_2)] - \log[f(\alpha_1)]\} + \{\log[B(s|\alpha_2)] - \log[B(s|\alpha_1)]\}$$

Definition 1 An agent is said to herd on the public belief (the expression in curly brackets on the extreme right-hand side) if her private distribution update (the left-hand side) is independent of her prior density (the first expression in curly brackets on the right-hand side).

Definition 2 If all agents herd, we have an informational cascade.

Our interest lies, naturally, in good signals. The strength of the priors of an optimist may be such that they outweigh a good signal on the impending demise of the project. There might be an informational cascade on a badly designed public works programme. In this case, when the left-hand side must be negative, with the first term on the right-hand side positive, the bureau of the government must get its numbers right so that the second term on the right-hand side is 'negative enough'. Since epsilon truth telling is a condition we impose on agents in the economics of asymmetric information, we see no reason not to extend it to the government. All that remains is to introduce a probability metric into our government bond issue process so it can function as this updating mechanism.

Now, the instantaneous spot rate contracted today, time t, r_{lt}, is riskless in that it is default free. Then the price of a default-free pure discount bond maturing at time t_1 and observed now at t is

$$B(t, t_1) = e^{-r_l(t,t_1)(t_1-t)} \tag{6}$$

Likewise,

$$B(t_1, t_2) = e^{-r_l(t_1,t_2)(t_2-t_1)} \tag{7}$$

Finally, and trivially,

$$B(t, t) \equiv B(T, T) \equiv 1 \tag{8}$$

Taking 6 + 7 = 8, we have

$$e^{-r_l(t,t_1)(t_1-t)} + e^{-r_l(t_1,t_2)(t_2-t_1)} = 1$$

To see the connection with our earlier notation, if the spot rates are constant in each interval,

$$e^{-r_{l1}} + e^{-r_{l2}} = 1$$

The task of the authorities, then, is then to set the interest rates so that either of the terms on the left-hand side equals x with the other term $1 - x$ and we have the bond price cum density function that can serve as the updating mechanism. Recall that the mechanism is supposed to work in the direction of truthful revelations, either generating informational cascades or supporting prior individual distributions.

Support of our framework can be culled from history. The Bank of France maintained its own discount rate and continuously violated the rules of the Gold Standard (Bazot et al. 2014). In 1908, the Banque announced its primary commitment to discounting commercial paper, ensuring appropriate liquidity levels in the support of national business. The Banque was a private institution so market principles were not foregone. Along with the profit objectives of commercial banks, other stakeholders were the State, merchants, and industrialists. The domestic asset portfolio played a pivotal role in smoothening external shocks and maintaining the stability of the discount rate. In short, international adjustments worked through sometimes massive short-term adjustments in the balance sheet of the Bank of France rather than through the discount rate. The possible trinity was not unlike ours: high foreign exchange reserves, a stable interest rate, and an expansion of domestic credit in response to international shocks.

4 Conclusion

We deal with the twin phenomena of depressed domestic economies and increasing international openness. The response has been to restore pride of place to output and growth at home and craft international arrangements that suit national concerns. A growing section of academic economists are beginning to entertain the possibility that classical neutrality and dichotomy results might not be applicable. Alternatives to the general equilibrium approach to these problems are being sought. Thus, money and finance must appear on the ground floor of models. We have discussed the efficacy of firewalls in this regard. Our approach is the input–output tradition that harks back to Quesnay. The dilemma is that in a closed world economy one country's expanded balance sheet might be another country's reduced balance sheet. On the one hand, 'foreign' assets or liabilities provide the 'home' country degrees of freedom in pursing own objectives. On the other, possible balance sheet crises might emerge out of 'foreign' bond holding, other things being equal. We offer a novel rewrite of the domestic bond as a contract written on tangible outcomes at home. The terms of the contract are effected by a class of banks. Profit-maximizing incentives are not impaired nor arbitrage conditions violated.

References

Bazot G, Bordo MD, Monnet E (2014) The price of stability: the balance sheet policy of the Banque de France and the gold standard (1880–1914), NBER Working Paper 20554, October 2014

Benes J, Kumhof M (2012) The Chicago plan revisited, IMF Working Paper 12/202, August 2012

Bordo M, James H (2015) Capital flows and domestic and international order: trilemmas from macro to political economy and, NBER Working Paper 21017, March 2015

Chamley CP (2004) Rational herds. Cambridge University Press, Cambridge

Combes J-L, Minen A, Sow, M (2013), Crises and exchange rate regimes: time to break down then bipolar view? Centre DÉtudes Et De Recherches Sur Le Developpment International, Etudes et Documents, p 26

Forrest JY-L (2014) Systems perspective on financial systems. CRC Press Taylor and Francis Group, London and New York

Gandolfo G (2002) International finance and open-economy macroeconomics. Springer, Vienna

Godley W, Lavoie M (2007) Monetary economics. Palgrave Macmillan, Basingstoke

Lavoie M (2013), Post-Keynesian monetary economics, Godley-like. In: Harcourt GK, Kriesler P (eds) The oxford handbook of post-Keynesian economics, vol 1: theory and origins, Oxford University Press, Oxford, pp 203–217